(2)

AD_____

CONTRACT NO: DAMD17-90-C-0050

TITLE: MOLECULAR STUDIES OF ALPHAVIRUS IMMUNOGENICITY

PRINCIPAL INVESTIGATOR: James H. Strauss, Ph.D.

CONTRACTING ORGANIZATION: California Institute of Technology
1201 East California Blvd.
Pasadena, California 91125

REPORT DATE: December 3, 1992

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21702-5012

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION California Institute of Technology 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) 1201 East California Blvd. Pasadena, California 91125	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract No. DAMD17-90-C-0050
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, Maryland 21702-5012	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102A; PROJECT NO. 301 61102BS12; TASK NO. AB; WORK UNIT ACCESSION NO. 117

11. TITLE (Include Security Classification)
Molecular Studies of Alphavirus Immunogenicity

12. PERSONAL AUTHOR(S)
James H. Strauss, Ph.D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 3/30/90 TO 9/30/92	1992 December 3	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Alphavirus/Sindbis virus/Whataroa virus/Aura virus/Ockelbo disease; BD, RAD IV; Immunogenetics; DNA; BL2; Biotechnology

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

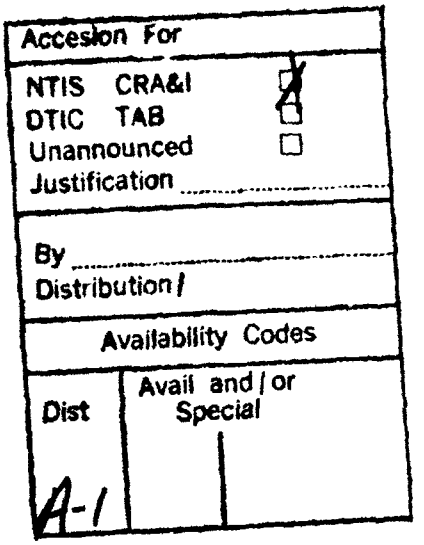
The alphaviruses consist of a group of 26 closely related viruses. Many of these viruses can cause disease in man, characterized by encephalitis, polyarthritis, feveror rash, depending upon the virus. In the 2.5 years of reasearch supported under this contract we have mapped antigenic epitopes in the structural glycoproteins of alphaviruses that lead to neutralization of virus infectivity upon reaction with and antibody, and have determined the sequence relationships of a number of Sindbis-like alphaviruses to one another and to other alphaviruses. We found that a domain of glycoprotein E2 of alphaviruses, between residues 170 and 220, was an important region for binding of monoclonal antibodies that neutralize virus infectivity, making it critical importance for the immune response required for protection from infection by the virus. In the determination of the relationships of alphaviruses to one another, we have determined complete or partial sequences of 8 different alphavirus RNAs. These include Ockelbo virus, a virus causing epidemic polyarthritis in northern Europe, strains of Sindbis virus from Africa, India, Australia and New Zealand and

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Virginia M. Miller	22b TELEPHONE (Include Area Code) 301-619-7325
22c. OFFICE SYMBOL SGRD-RMI-S	

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

19. ABSTRACT Continued

Aura virus from South America. We found that the Sindbis-like viruses possess certain key features in common and are all closely related to one another. Aura virus in the first true representative of the Sindbis viruses found in the Americas and demonstrates that these viruses are global in their distribution. We have also developed improved methods for rapidly sequencing large viral RNA genomes.



Contract No. DAMD17-90-0050

Final Report

List of personnel receiving pay, effort percentage and graduate degrees resulting from the contract support.

James H. Strauss, Professor, Principal Investigator 10% Effort 3/30/90-9/30/92

Richard Kuhn, Senior Research 50% Effort 4/1/90-12/31/91

Kangsheng Wang,Graduate Student 50% 4/1/90-6/30/91
Kangsheng Wang,Research Fellow 50% Effort 7/1/91-5/31/92
completed Ph.D requirements 5/91

Hong Zhang Research Assist. 75% Effort 4/23/90-8/23/92

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

JKS In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

12-3-92

PI - Signature DATE

TABLE OF CONTENTS

INTRODUCTION

The alphaviruses are a widespread group of human pathogens that are present virtually everywhere in the world (Griffin, 1986; Monath, 1988; Peters and Dalrymple, 1990) They are mosquito-borne viruses and thus are particularly prevalent in tropical areas where mosquitoes abound and problems of overwintering by the virus do not arise, but are also present in temperate areas of the world including the United States. They have the capacity to replicate in the mosquito vector as well as in human hosts or in various species of birds and mammals. Old World alphaviruses are, in general, capable of causing a painful and disabling disease in man characterized by fever, rash and arthralgia. In the cases of the Ockelbo strain of Sindbis virus and of Ross River virus, this arthralgia manifests as a polyarthritis that may in some cases last for months or years. Many of the New World alphaviruses can cause fatal encephalitis in man. Our program attempts to understand the molecular basis of alphavirus immunogenicity and to determine the relationships of alphaviruses to one another, and has developed in collaboration with Drs. Alan Schmaljohn and Joel Dalrymple of USAMRIID.

METHODS USED

Virus Strains. Viruses used in this study were from the collection of Dr. J. M. Dalrymple of USAMRIID. Viruses were grown in BHK cells, in secondary chicken embryo fibroblast cells, or in mosquito cells, purified, and RNA prepared as described (Ou et al., 1981; Shirako et al., 1991).

cDNA Clones. cDNA clones were made in one of two ways. The first method used standard procedures in which first strand cDNA was made using oligo(dT) as primer and second strand synthesis was by the method of Gubler and Hoffman (Gubler and Hoffman, 1983; Sambrook et al., 1989). These cloning methods, as well as the methods of DNA sequencing and RNA sequencing, have been described in numerous publications from our laboratory over the years (Hahn et al., 1985; Rice et al., 1985; Rice and Strauss, 1981; Shirako et al., 1991; Strauss et al., 1984).

In a second approach, we developed methods suitable for high throughput automated DNA sequencing, in order to speed up the acquisition of sequence data. Whataroa virus was chosen as a test virus. The methods were described in detail in our annual report of 4/23/92. Briefly, first strand cDNA synthesis used random priming and second strand cDNA was synthesized by the method of Gubler and Hoffman (Gubler and Hoffman, 1983). After blunt-ending the double-stranded cDNA, the internal *Eco*RI restriction sites were methylated and the DNA was electrophoresed in an agarose gel. *Eco*RI linkers were attached to the 2-4 kb fraction and the DNA cloned in the *Eco*RI site of a suitable vector. One hundred clones that resulted from this cloning were characterized by restriction analysis and many of them were sequenced using an Applied Biosystems automated DNA sequencer.

Construction and Screening of the Bacteriophage Lambda Library. Sindbis virus strain AR339 from A. Schmaljohn at USAMRIID was grown in monolayers of primary chicken cells (Pierce et al., 1974). Virus was purified as described (Bell et al., 1978), disrupted with 0.5% SDS, and 49S genomic RNA extracted with phenol/chloroform (Hsu et al., 1973). After two ethanol precipitations, RNA was suspended in distilled water and stored at -70°C until used as a template for cDNA

synthesis. A λgt11 expression library containing short inserts of Sindbis cDNA was constructed by a modification of the procedure of Young and Davis (Young and Davis, 1983). cDNA synthesis was randomly primed with sonicated salmon testis DNA. After flush-ending the product with the Klenow fragement of DNA polymerase I, methylation with *Eco*RI methytransferase, and addition of *Eco*RI linkers, the modified cDNA was digested with an excess of *Eco*RI restriction enzyme. The digested DNA was fractionated on a Sephadex CL-6B column, and cDNA fragments 100-300 base pairs in size were pooled and ligated to dephosphorylated λgt11 arms (Promega). After in vitro packaging into phage heads (Stratagene), phage plaques were grown for 6 h at 42°C. Nitrocellulose disks soaked in 10 mM isopropyl thio-β-D-galactopyranoside were then placed on top of the agar layer, and the plates were transferred to 37°C for 15 h. The filters were lifted and washed successively in 10 mM Tris-Cl pH 7.5 and 150 mM NaCl containing 5% nonfat milk. The filters were incubated overnight at 4°C with a monoclonal antibody in PBS solution containing 5% nonfat milk, washed, and the filters were then incubated at least two hours at room temperature in the presence of ^{125}I-conjugated protein G (0.5 μCi/ml in 5% nonfat milk). After washing and drying, the filters were exposed overnight at -80°C to Kodak-X-Omat film. Immunoreactive phage were picked and rescreened until a uniformly reactive population was obtained.

MAPPING OF NEUTRALIZING ANTIGENIC EPITOPES OF ALPHAVIRUSES

We have localized a site in alphavirus glycoprotein E2 that binds neutralizing antibodies. Characterization of such immunogenic domains is important in developing vaccines, because neutralizing antibodies are thought to be particularly important in protecting a vaccinee from viral infection. We

developed a novel approach in which λgt11 expression libraries were constructed that expressed parts of the Sindbis genome, and these were screened with neutralizing monoclonal antibodies (MAbs). Many neutralizing antibodies react with discontinuous epitopes and thus will not react with a chimeric protein expressed in a λgt11 library. However, we succeeded in identifying one MAb which bound to specific clones within the λgt11 library (Wang and Strauss, 1991). Four λgt11 clones were found that reacted with MAb23, and a schematic of these four clones in relation to the Sindbis virus genome is shown in Fig. 1. The four clones all contain overlapping inserts from the E2 region of the genome, and the sequence of E2 from residues 173 to 220 is present in all. This demonstrates directly that this neutralizing MAb binds to glycoprotein E2 of Sindbis virus between residues 173 and 220.

The result with MAb23 confirmed and extended results in which variants of the virus selected to be resistant to neutralizing MAbs were sequenced in order to identify the regions within the glycoproteins of the virus with which the antibodies react (Strauss et al., 1991). This is illustrated in Fig. 1 in which the sequence of E2 between residues 173 and 220 is shown, and the location of many variants that render the virus resistant to neutralization by several MAbs is indicated. It is clear that the domain between residues 170 and 220 of glycoprotein E2 of alphaviruses is particularly important for the antibody response of a host. We have estimated that 90% of the neutralizing antibodies produced by an infected mouse are directed against this E2 domain (Strauss et al., 1991)

This domain of E2 identified as being important for reactivity with neutralizing antibodies also appears to be important for virus attachment to host cell receptors. First, many neutralizing antibodies are thought to inactivate the virus by binding to the domain that interacts with the cell receptor, thus blocking virus binding to the cell. Second, antiidiotypic antibodies made to MAbs that bind

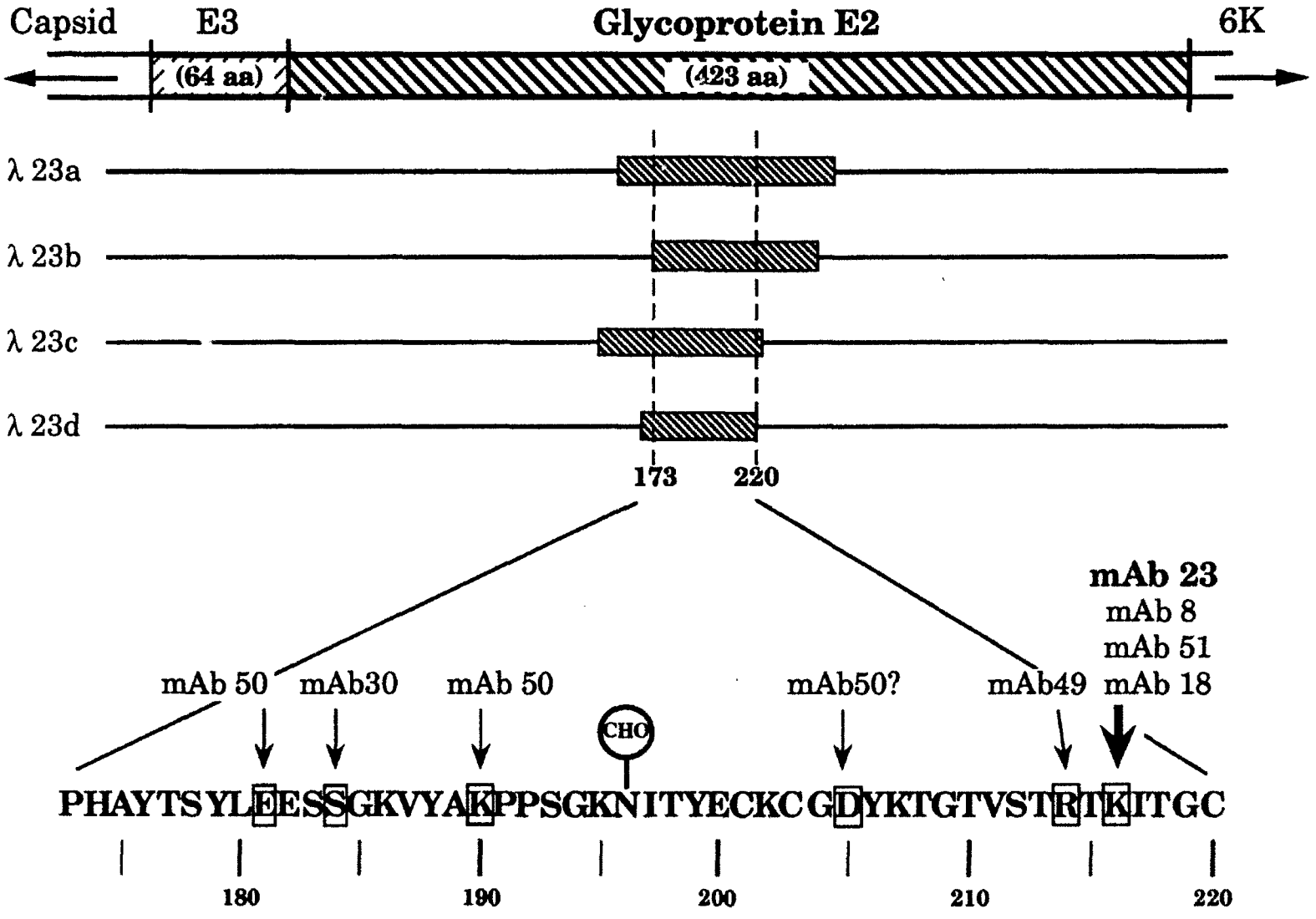


Figure 1. Schematic representation of an antigenically important domain of Sindbis virus glycoprotein E2. The relative locations of the inserts in four λgt11 clones reactive with MAb 23 are mapped. The overlap region in these four clones between residues 173 and 220 of E2 is expanded below, with a number of key features indicated. Residues altered in variants resistant to MAbs are boxed and a carbohydrate attachment site is indicated with a stalked symbol (CHO).

to this domain of Sindbis E2 function as antireceptor antibodies (Wang et al., 1991). Third, changes in this region of E2 alter the ability of the virus to bind to neuronal cells (Ubol and Griffin, 1991). The simplest interpretation of these results is that the E2 domain between 170 and 220 binds to a cell receptor to initiate infection.

ALPHAVIRUSES EXAMINED FOR SEQUENCE RELATIONSHIPS

We have examined 12 strains of alphaviruses for their relationships to one another. These 12 viruses are shown in Fig. 2 together with the source from which they were isolated, their year of isolation, and the location in which they were isolated. Strains to be examined were chosen in consultation with Dr. Joel Dalrymple of USAMRIID, and were chosen on the basis of geography, year of isolation, potential for human disease, and, in the case of Aura virus, as a possible parent for the emergent virus Western equine encephalitis virus.

SEQUENCE ANALYSIS OF OCKELBO VIRUS

We have determined the complete nucleotide sequence of the genome of Ockelbo virus. This virus was chosen for analysis because it causes epidemics of polyarthritis in humans, a disabling disease that can last for months. The sequence of the virus isolated in 1982 in Edsbyn, Sweden, is shown in Fig. 3. The viral genome is 11,708 nucleotides in length excluding the poly(A) tail. The genome is identical in organization to that of the Sindbis virus AR339 strain (Strauss et al., 1984) isolated in Sindbis, Egypt in 1952 (Taylor et al., 1955). There are only 672 nucleotide differences between the two viruses (5.7% divergence) that result in 97 amino acid changes (2.6) divergence. Thus more that 85% of all

Name	Strain	Source	Year	Location	Reference
Subgroup I					
Sindbis	AR339	Mosquito *(Culex univittatus)*	1952	Egypt	Taylor et al., (1955)
Sindbis	MP684	Mosquito *(Mansonia fuscopennata)*	1958	Uganda	
Sindbis	Girdwood	Human	1963	South Africa	Malherbe et al., (1963)
Sindbis	R33	Reed Warbler *(Acrocephalus scirpaceous)*	1971	Czechoslovakia	
Sindbis	1038	Turtle Dove *(Streptopelia turtur)*	1964	Israel	
Ockelbo	Edsbyn 82-5	Mosquito pool (*Culiseta* spp.)	1982	Edsbyn village, Sweden	Niklasson et al., (1984)
Ockelbo	Edsbyn 83M107	Mosquito *(Culiseta morsitans)*	1983	Edsbyn village, Sweden	
Karelian Fever	LEIV 9298	Mosquito *(Aedes communis)*	1983	Central Karelia, USSR	Lvov et al., (1984, 1988)
Subgroup II					
Sindbis	MM2215	Mosquito *(Culex tritaeniorhynchus)*	1955	Indonesia	
Sindbis	A-1036	Mite *(Bdellonyssus bursa)*	1953	India	Shah et al., (1960)
Sindbis	MRM18520	Mosquito (unidentified)	1975	Queensland, Australia	
Subgroup III					
Whataroa	M78	Mosquito pool	1962	New Zealand	
Subgroup IV					
Aura	AR10315	Mosquito *(Culex* spp.)	1959	Brazil	Causey et al. 1963

Figure 2 . Strains of Sindbis virus and related viruses used in this study.

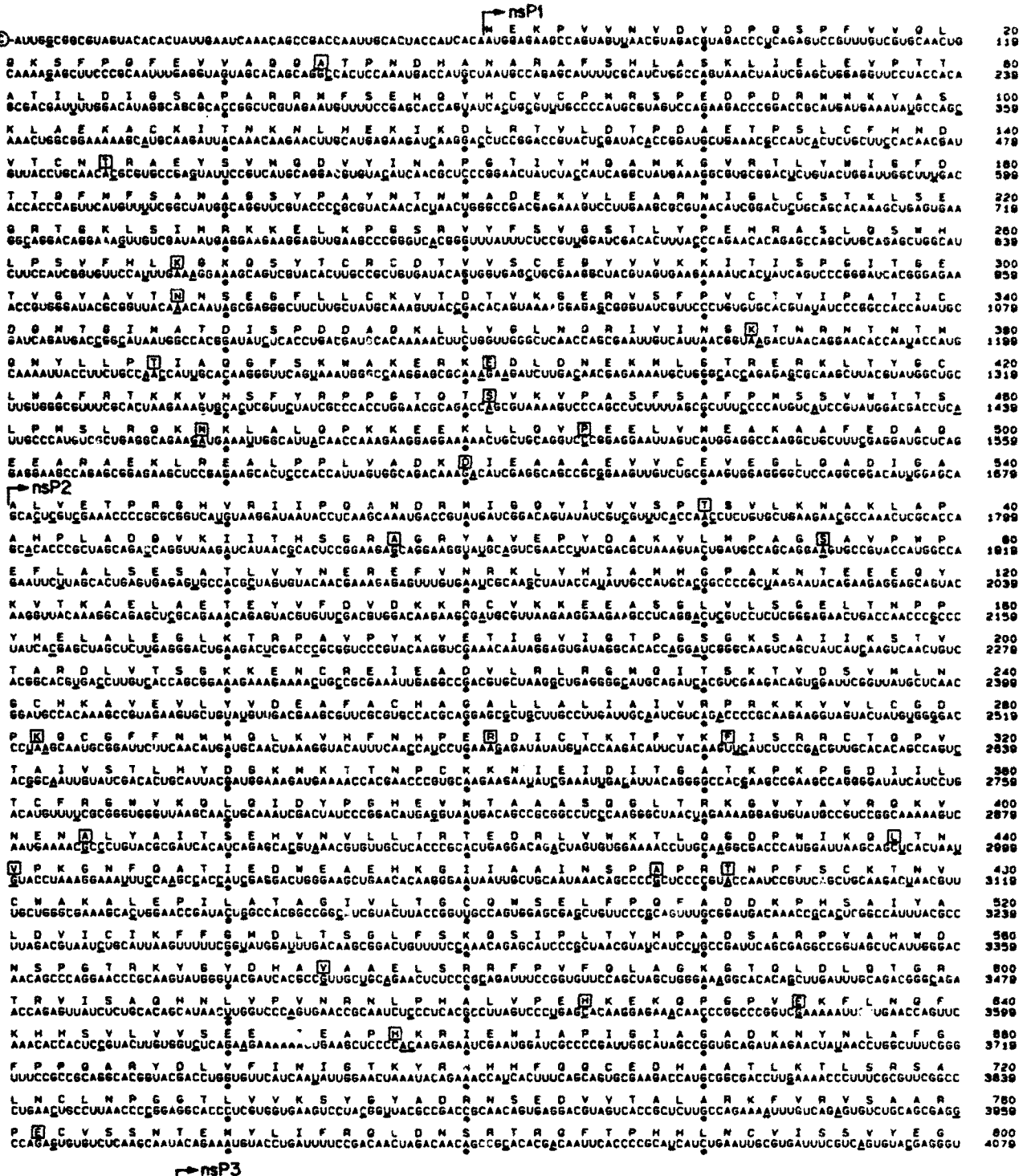


Figure 3a. See legend on last page of this sequence.

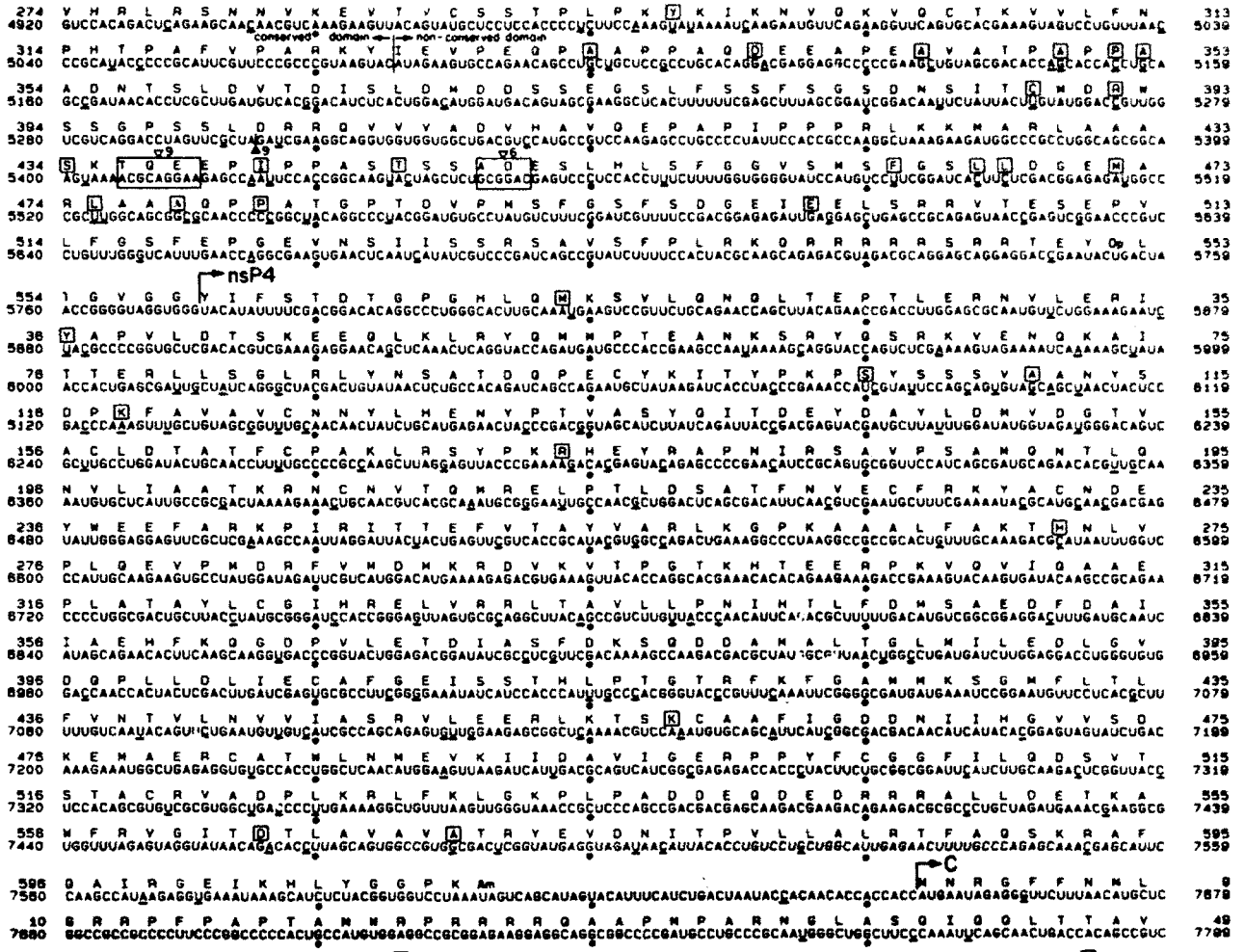


Figure 3b. See legend on the last page of this sequence.

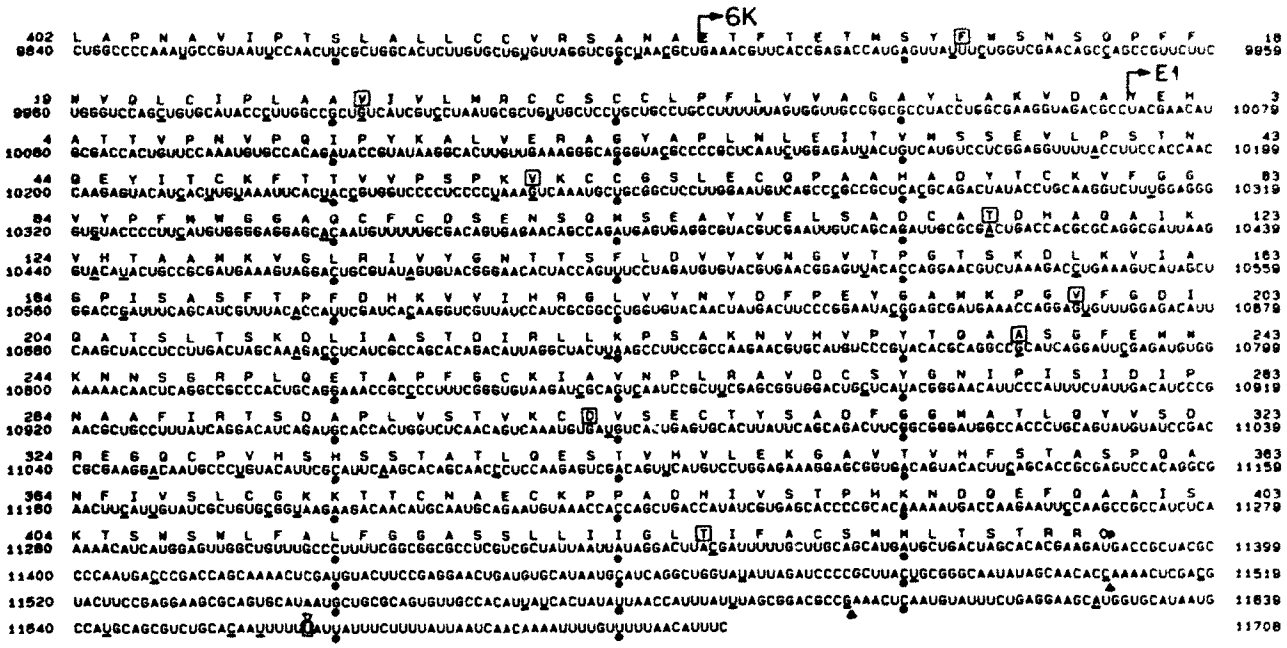


Figure 3c. Complete nucleotide sequence of the Ockelbo virus genome. The sequence is shown from 5' to 3' and translated using the single letter amino acid code. Nucleotides different from those in HRSP are underlined, and changed amino acids are boxed. Deletions relative to HR are indicated by solid triangles pointing upward and the number of residues deleted. Insertions have both amino acids and nucleotides boxed together, and an open triangle pointing downward. Termination codons are labelled Am (Amber, UAG) or Op (Opal, UGA) as appropriate. Nucleotides are numbered 5' to 3'; amino acid numbering begins again at the beginning of each final protein product.

nucleotide differences are silent, illustrating the importance of conservation of amino acid sequence.

Glycoprotein E2 is particularly important for antigenicity, as described above, and changes in E2 have been associated with changes in virulence (Lustig et al., 1988; Olmsted et al., 1986; Strauss et al., 1991; Tucker and Griffin, 1991). The differences in glycoprotein E2 between six strains of Sindbis virus are listed in Fig. 4. The residues at positions 172, 209, 212, and 216 are known to be important determinants of the antigenicity of the virus (Strauss et al., 1991), and the changes in these positions are important for the differences in the cross-reactivity of the viruses with antibodies. The residues at 55 and 172 are known to be important determinants of the neurovirulence of the virus in a mouse model (Lustig et al., 1988), and it is possible that the amino acid difference at position 55 may be important for the increased virulence of Ockelbo virus compared to the other strains of Sindbis virus in Fig. 4.

ANALYSIS OF 3' TERMINAL NONTRANSLATED SEQUENCE

To study the relationships among a number of Sindbis viruses present in nature, the sequences of the 3' nontranslated regions (NTR) were obtained for a number of strains. These sequences are shown in Fig. 5. The sequence identity throughout this region is greater than 80% for all viruses shown, and the sequence organization is identical except for a few scattered insertions and deletions. In the 3' NTR there are three repeated elements that are highly conserved (boxed in the figure). As an example of the conservation of these elements, there are 49 differences in the 3' NTRs of the Australian and AR339 strains that occur outside the repeated elements (24.1% divergence) but only 7 changes within these elements (5.8% divergence), and the overall divergence is

RESIDUE	AR339				S.A. AR86	OCKELBO
	HRSP	DG	AS	RJ		
1	S	S	S	R	S	S
3	I	T	T	T	T	T
23	V	E	E	E	E	E
29	V	V	V	V	I	I
55	Q	Q	Q	Q	Q	K
61	A	A	A	A	S	T
69	L	L	L	L	L	F
70	K	K	E	E	E	E
116	V	V	V	V	A	A
126	L	L	L	L	M	M
172	R	G	G	G	G	G
209	G	R	G	G	G	G
212	S	S	S	S	T	T
216	E	E	K	E	E	E
243	L	L	L	L	S	L
247	D	D	D	D	A	A
277	I	I	I	I	V	I
312	V	V	V	V	I	I
375	T	T	T	T	A	A
386	V	V	V	V	A	A

Figure 4. Amino acid differences in the glycoprotein E2 of various Sindbis strains. The sequence of HRSP is from Strauss et al. (1984); The sequence marked DG is the SV1A strain published in Lustig et al.(1988). AS is our unpublished sequence of the strain used by A. Schmaljohn for the isolation of antigenic escape mutants (Stec. et al., 1986); RJ is the sequence from Davis et al. (1986) of a laboratory strain from Robert Johnston. The sequence of AR86 was reported in Russell et al. (1989), and the Ockelbo sequence was presented in Figure 3.

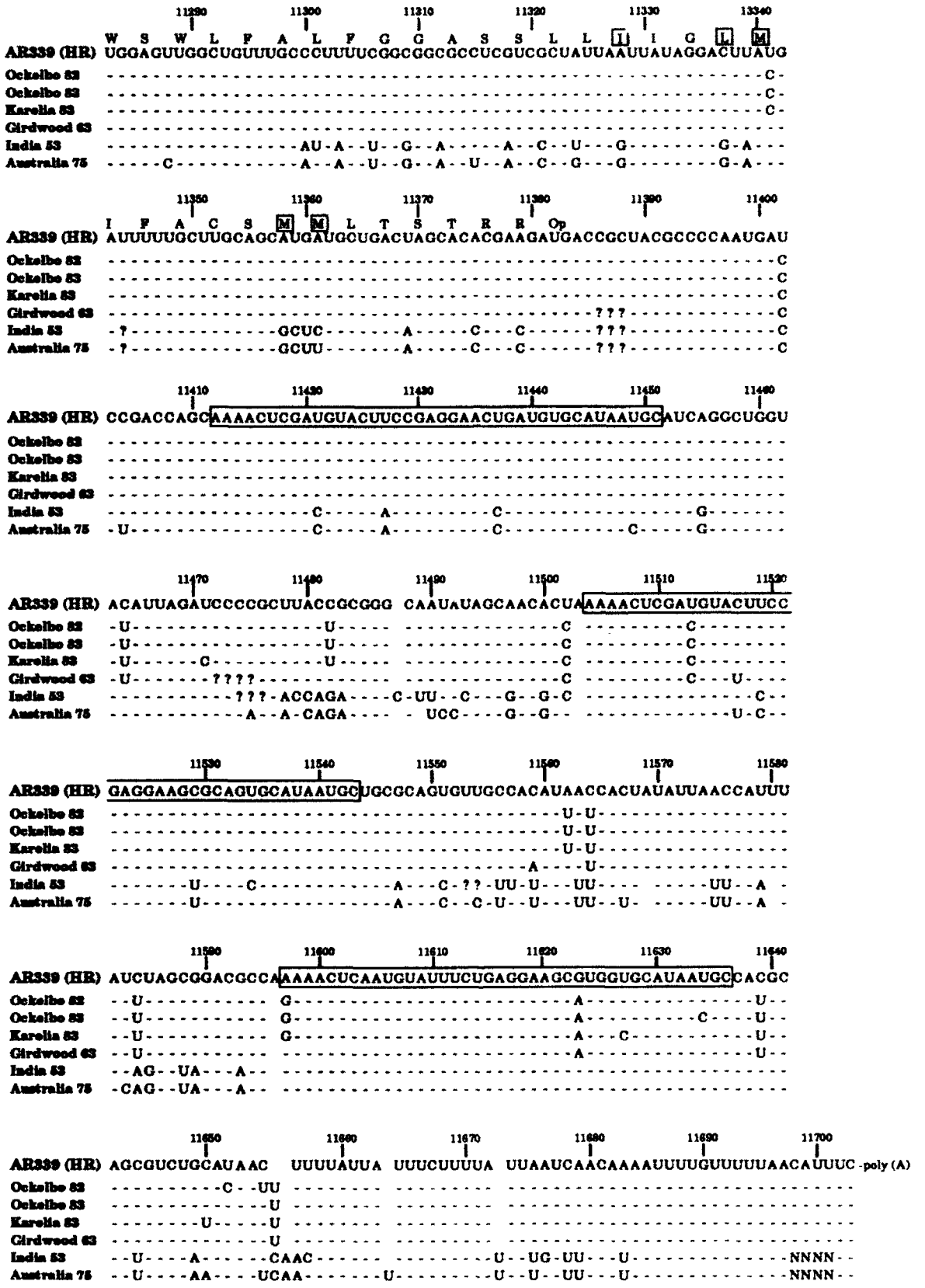


Figure 5. Sequence of the 3' termini of several Sindbis viruses. The sequences of Ockelbo 83M107, Karelian fever, and South African Sindbis (Girdwood) were determined from cloned cDNA. Those of the Indian A1036 and Australian 18520 isolates were determined directly from RNA by dideoxy sequencing using reverse transcriptase and a T12GA primer. The Ockelbo 82 sequence is from Fig. 3 and that of AR339 (HRSP) is from Strauss et al. (1984). Three repeated sequence elements of 40 nucleotides are boxed. The translated sequence is for AR339 (HRSP) and any amino acid that differs in the other viruses is boxed. This figure is from Shirako et al. (1991).

18.1%. From such analysis, we propose that these repeated and conserved elements play an important role in viral RNA replication, and this role is probably more important in mosquito cells than in vertebrate cells (Kuhn et al., 1990).

The relationships among these viruses is illustrated in Fig. 6. Three points are obvious from this diagram. One is that the Sindbis strains analyzed can be divided into a European-African group and an Asian-Australian group. The second point is that Ockelbo virus and Karelian fever virus are virtually identical. The third point is that Ockelbo virus is more closely related to the South African strain of Sindbis virus isolated in 1963 (and which is also capable of causing human illness) than it is to the Egyptian strain isolated in 1952. We conclude from this last point that Ockelbo virus was probably introduced into Northern Sweden from South Africa in the 1960s, from where it spread into Finland (where it causes the disease called Pogosta) and the Karelian region of Russia.

SEQUENCE STUDIES OF AURA RNA

We have obtained the sequence of essentially all of the genome of Aura virus and are currently assembling this sequence. We were particularly interested in this virus because we have previously shown that Western equine encephalitis virus (WEE), previously thought to be closely related to Sindbis virus, is in fact a recombinant virus in which most of the genome was derived from Eastern equine encephalitis virus and only the surface glycoproteins were derived from a Sindbis-like virus (Hahn et al., 1988). Thus the question arose as to whether there is a virus found in the Americas that is closely related to Sindbis and that could have served as the second parent of WEE. The question is of particular interest because WEE emerged from a recombination event.

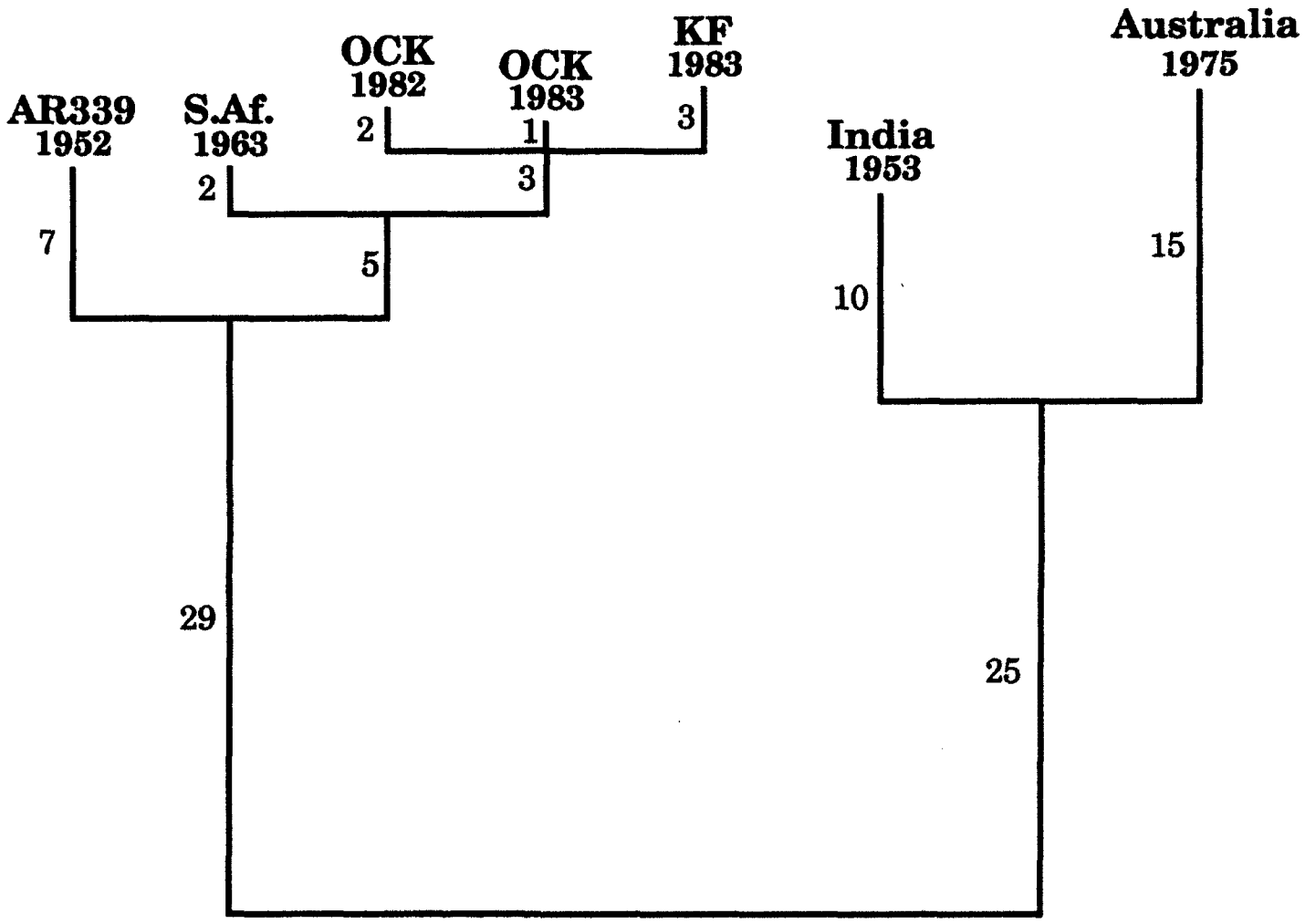


Figure 6 . Evolutionary tree of strains of Sindbis virus. The vertical distances indicate the number of nucleotide differences between any two strains in the 3'terminal 420 nucleotides. The horizontal distances are arbitrary. The tree was constructed from a difference matrix by iteration to give the best possible representation with the minimum number of branch points. Nucleotide differences between any two strains can be calculated by summing the numbers on the vertical branches connecting the two strains to be compared.

The sequence of about 5000 nucleotides of Aura RNA in the nonstructural protein coding region is shown in Fig. 7. This sequence begins in the 5' NTR and continues through nsP1, nsP2, and part of nsP3. From this sequence, it is clear that Aura virus is closely related to Sindbis virus. Comparison of the amino acid sequences of Sindbis virus and of Aura virus in the region represented by the Aura sequence in Fig. 7 shows that the two sequences are 80% identical, illustrating that Aura is in fact a Sindbis-like virus. We also found that the 3' NTR of Aura RNA is Sindbis-like. As described above, Sindbis-like viruses contain three copies of a conserved sequence element that we postulate is important for RNA replication. Although other alphaviruses often contain repeated sequence elements, these elements are completely different in sequence from the Sindbis sequence. Furthermore, WEE lacks the characteristic Sindbis 3' NTR, and contains instead a chimeric 3' NTR. Thus Aura virus represents the first known example of a true Sindbis-like virus in the Americas

Aura virus is widely distributed in South America, having been isolated in Brazil and in Northern Argentina. Analysis of the data is not yet complete, but it is possible that Aura virus represents the ancestral Sindbis-like virus, and that it was transmitted to the Old World to serve as the founder of the Sindbis viruses in the Old World, as we previously postulated (Levinson et al., 1990). Aura virus may have served as one of the parents of WEE, contributing its glycoproteins to this recombinant virus (Hahn et al., 1988).

```
1           ACT AGT ACT TGT ACT ACA GAA TTA ACT GCC GTG TGC CGC CCG CTA AAC TAG CCC CAA TCA

61          TCG AAA ATG GAG AAA CCG ACA GTG CAC GTT GAC GTA GAC CCC CAA AGT CCG TTT GTG CTA
                    met glu lys pro thr val his val asp val asp pro gln ser pro phe val leu

121/19      CAA CTG CAG AAG AGT TTC CCA CAA TTC GAG ATT GTG GCT CAG CAG GTC ACT CCG AAT GAC
            gln leu gln lys ser phe pro gln phe glu ile val ala gln gln val thr pro asn asp

181/39      CAT GCT AAT GCC AGA GCT TTT TCG CAT CTG GCT AGT AAA CTG ATC GAA CAT GAG ATC CCC
            his ala asn ala arg ala phe ser his leu ala ser lys leu ile glu his glu ile pro

241/59      ACC TCA GTT ACG ATC TTG GAC ATA GGA AGC GCA CCA GCT CGT AGA ATG TAT TCC GAG CAT
            thr ser val thr ile leu asp ile gly ser ala pro ala arg arg met tyr ser glu his

301/79      AAG TAT CAC TGT GTG TGC CCC ATG CGT AGT CCT GAA GAC CCG GAC CGT CTT ATG AAT TAC
            lys tyr his cys val cys pro met arg ser pro glu asp pro asp arg leu met asn tyr

361/99      GCA TCC CGA CTC GCA GAC AAA GCA GGG GAA ATT ACC AAC AAG AGG CTG CAT GAT AAA CTT
            ala ser arg leu ala asp lys ala gly glu ile thr asn lys arg leu his asp lys leu

421/119     GCA GAC CTC AAG TCG GTC CTC GAG TCG CCG GAT GCT GAA ACT GGT ACC ATT TGT TTC CAC
            ala asp leu lys ser val leu glu ser pro asp ala glu thr gly thr ile cys phe his

481/139     AAT GAC GTA ATA TGC CGT ACG ACA GCG GAG GTA TCA GTT ATG CAA AAT GTG TAT ATC AAT
            asn asp val ile cys arg thr thr ala glu val ser val met gln asn val tyr ile asn

541/159     GCA CCT TCG ACC ATT TAC CAT CAG GCC CTA AAG GGA GTC AGA AAA CTG TAT TGG ATC GGG
            ala pro ser thr ile tyr his gln ala leu lys gly val arg lys leu tyr trp ile gly

601/179     TTC GAT ACA ACG CAG TTT ATG TTC TCC TCG ATG GCA GGG TCG TAT CCG TCC TAC AAT ACT
            phe asp thr thr gln phe met phe ser ser met ala gly ser tyr pro ser tyr asn thr

661/199     AAT TGG GCC GAT GAA AGG GTG CTG GAA GCG CGT AAT ATA GGC CTA TGT AGC ACG AAG CTG
            asn trp ala asp glu arg val leu glu ala arg asn ile gly leu cys ser thr lys leu

721/219     AGA GAG GGT ACG ATG GGC AAA CTG TCT ACC TTC CGG AAA AAG GCC TTG AAA CCT GGA ACT
            arg glu gly thr met gly lys leu ser thr phe arg lys lys ala leu lys pro gly thr

781/239     AAC GTG TAC TTC TCT GTC GGT TCG ACA CTC TAC CCT GAG AAT AGA GCG GAC CTG CAG AGT
            asn val tyr phe ser val gly ser thr leu tyr pro glu asn arg ala asp leu gln ser

841/259     TGG CAC CTA CCA TCT GTG TTC CAC TTG AAA GGT AAA CAA TCC TTT ACG TGC CGC TGT GAT
            trp his leu pro ser val phe his leu lys gly lys gln ser phe thr cys arg cys asp

901/279     ACG GCG GTT AAC TGC GAA GGA TAC GTA GTC AAG AAG ATC ACC ATC AGC CCC GGG ATC ACG
            thr ala val asn cys glu gly tyr val val lys lys ile thr ile ser pro gly ile thr

961/299     GGG CGT GTC AAT CGG TAC ACT GTG ACT AAC AAC AGC GAG GGA TTC TTG CTG TGT AAG ATC
            gly arg val asn arg tyr thr val thr asn asn ser glu gly phe leu leu cys lys ile

1021/319    ACA GAT ACG GTC AAA GGG GAG CGT GTA TCG TTC CCT GTC TGT ACG TAT ATT CCA CCT TCA
            thr asp thr val lys gly glu arg val ser phe pro val cys thr tyr ile pro pro ser

1081/339    ATC TGT GAC CAA ATG ACA GGT ATA TTG GCC ACT GAT ATC CAA CCC GAA GAC GCG CAA AAG
            ile cys asp gln met thr gly ile leu ala thr asp ile gln pro glu asp ala gln lys
```

Figure 7a. See legend on last page of this sequence

```
1141/359   TTG CTG GTA GGA CTG AAC CAA CGC ATA GTC GTG AAC GGA AAA ACT AAT AGA AAC ACC AAC
           leu leu val gly leu asn gln arg ile val val asn gly lys thr asn arg asn thr asn

1201/379   ACG ATG CAG AAC TAT CTC CTG CCC GCG GTG GCT ACA GGT CTG AGT AAA TGG GCC AAA GAA
           thr met gln asn tyr leu leu pro ala val ala thr gly leu ser lys trp ala lys glu

1261/399   AGA AAG GCA GAC TGC AGT GAC GAG AAA CCA TTG AAT GTG AGA GAA CGC AAA CTA GCT TTC
           arg lys ala asp cys ser asp glu lys pro leu asn val arg glu arg lys leu ala phe

1321/419   GGT TGC CTA TGG GCT TTC AAG ACC AAG AAG ATC CAT TCT TTT TAC CGC CCG CCA GGC ACG
           gly cys leu trp ala phe lys thr lys lys ile his ser phe tyr arg pro pro gly thr

1381/439   CAG ACT ATA GTA AAA GTC GCA GCG GAA TTC AGT GCG TTC CCT ATG TCC TCG GTG TGG ACT
           gln thr ile val lys val ala ala glu phe ser ala phe pro met ser ser val trp thr

1441/459   ACG TCA CTG CCA ATG TCA CTG AGA CAG AAA GTT AAA CTG CTT CTT GTA AAG AAA ACC AAT
           thr ser leu pro met ser leu arg gln lys val lys leu leu leu val lys lys thr asn

1501/479   AAA CCG GTA GTC ACT ATT ACT GAC ACT GCG GTA AAA AAC GCA CAA GAG GCA TAT AAC GAA
           lys pro val val thr ile thr asp thr ala val lys asn ala gln glu ala tyr asn glu

1561/499   GCC GTC GAG ACA GCA GAA GCG GAG GAG AAA GCG AAG GCC TTA CCT CCG CTG AAG CCG ACG
           ala val glu thr ala glu ala glu glu lys ala lys ala leu pro pro leu lys pro thr

1621/519   GCA CCC CCT GTA GCG GAG GAC GTC AAA TGC GAG GTC ACC GAC CTG GTA GAC GAT GCG GGA
           ala pro pro val ala glu asp val lys cys glu val thr asp leu val asp asp ala gly

1681/539   GCG GCC CTG GTC GAG ACG CCC CGG GGA AAG ATA AAA ATT ATC CCA CAG GAA GGG GAC GTG
           ala ala leu val glu thr pro arg gly lys ile lys ile ile pro gln glu gly asp val

1741/559   CGT ATT GGT TCC TAC ACA GTC ATT TCT CCA GCG GCA GTC CTT AGA AAT CAA CAA CTG GAG
           arg ile gly ser tyr thr val ile ser pro ala ala val leu arg asn gln gln leu glu

1801/579   CCA ATC CAC GAG TTA GCA GAG CAG GTG AAA ATT ATC ACG CAC GGT GGC CGA ACA GGC AGG
           pro ile his glu leu ala glu gln val lys ile ile thr his gly gly arg thr gly arg

1861/599   TAT TCC GTC GAA CCT TAC GAT GCT AAG GTT CTC CTG CCA ACA GGA TGC CCC ATG TCC TGG
           tyr ser val glu pro tyr asp ala lys val leu leu pro thr gly cys pro met ser trp

1921/619   CAA CAT TTC GCG GCC TTG AGC GAA AGC GCT ACG TTA GTC TAC AAT GAG AGA GAG TTC CTG
           gln his phe ala ala leu ser glu ser ala thr leu val tyr asn glu arg glu phe leu

1981/639   AAC CGG AAA CTC CAT CAC ATC GCT ACG AAG GGT GCG GCA AAA AAC ACT GAG GAA GAA CAA
           asn arg lys leu his his ile ala thr lys gly ala ala lys asn thr glu glu glu gln

2041/659   TAC AAA GTA TGC AAA GCT AAA GAC ACG GAT CAT GAG TAC GTA TAC GAC GTA GAT GCC AGA
           tyr lys val cys lys ala lys asp thr asp his glu tyr val tyr asp val asp ala arg

2101/679   AAA TGC GTA AAA AGA GAG CAT GCA CAA GGG CTA GTA CTA GTT GGG GAA CTA ACT AAT CCG
           lys cys val lys arg glu his ala gln gly leu val leu val gly glu leu thr asn pro

2161/699   CCT TAC CAC GAG CTG GCA TAC GAA GGA TTA CGT ACA CGA CCC GCT GCC CCT TAC CAT ATC
           pro tyr his glu leu ala tyr glu gly leu arg thr arg pro ala ala pro tyr his ile
```

Figure 7b. See legend on last page of this sequence

2221/719 GAA ACA CTG GGG GTC ATT GGA ACA CCG GGG TCA GGT AAG TCG GCC ATC ATA AAA TCT ACG
glu thr leu gly val ile gly thr pro gly ser gly lys ser ala ile ile lys ser thr

2281/739 GTA ACA CTA AAA GAC CTC GTA ACT AGC GGT AAG AAA GAA AAT TGC AAA GAA ATA GAG AAT
val thr leu lys asp leu val thr ser gly lys lys glu asn cys lys glu ile glu asn

2341/759 GAC GTC CAG AAA ATG CGG GGA ATG ACT ATA GCT ACG AGA ACG GTA GAC TCG GTA CTT CTT
asp val gln lys met arg gly met thr ile ala thr arg thr val asp ser val leu leu

2401/779 AAT GGA TGG AAG AAA GCA GTA GAC GTC CTA TAT GTG GAT GAA GCG TTT GCA TGT CAT GCA
asn gly trp lys lys ala val asp val leu tyr val asp glu ala phe ala cys his ala

2461/799 GGC ACC TTA ATG GCA TTG ATT GCC ATT GTC AAA CCG AGA CGT AAA GTA GTA CTG TGC GGC
gly thr leu met ala leu ile ala ile val lys pro arg arg lys val val leu cys gly

2521/819 GAC CCG AAG CAG TGG CCC TTC TTT AAT TTA ATG CAA CTG AAG GTA AAC TTC AAC AAC CCC
asp pro lys gln trp pro phe phe asn leu met gln leu lys val asn phe asn asn pro

2581/839 GAG CGA GAC CTG TGT ACT TCC ACC CAT TAT AAA TAT ATC TCT CGC AGG TGC ACC CAA CCT
glu arg asp leu cys thr ser thr his tyr lys tyr ile ser arg arg cys thr gln pro

2641/859 GTT ACA GCC ATA GTG TCT ACA TTA CAC TAT GAC GGA AAG ATG AGG ACT ACG AAT CCC TGC
val thr ala ile val ser thr leu his tyr asp gly lys met arg thr thr asn pro cys

2701/879 AAA AGG GCT ATC GAA ATA GAC GTA AAC GGA TCG ACT AAG CCC AAG AAA GGA GAC ATA GTG
lys arg ala ile glu ile asp val asn gly ser thr lys pro lys lys gly asp ile val

2761/899 TTG ACG TGT TTC CGT GGG TGG GTT AAG CAG GGG CAA ATC GAT TAC CCC GGA CCC GGA GGT
leu thr cys phe arg gly trp val lys gln gly gln ile asp tyr pro gly pro gly gly

2821/919 CAT GAC CGT GCA GCT TCT CAA GGG CTA ACC AGA AGG GGC GTT TAT GCG GTC AGA CAG AAA
his asp arg ala ala ser gln gly leu thr arg arg gly val tyr ala val arg gln lys

2881/939 GTA AAT GAA AAC CCA CTA TAT GCA GAG AAG TCA GAA CAC GTT AAC GTG TTA CTT ACT AGG
val asn glu asn pro leu tyr ala glu lys ser glu his val asn val leu leu thr arg

2941/959 ACG GAA GAT CGC ATA GTG TGG AAG ACA CTG CAA GGG GAT CCT TGG ATT AAG TAC CTC ACT
thr glu asp arg ile val trp lys thr leu gln gly asp pro trp ile lys tyr leu thr

3001/979 AAC GTT CCA AAA GGG AAC TTT ACA GCC ACT TTA GAA GAA TGG CAG GCG GAA CAC GAG GAC
asn val pro lys gly asn phe thr ala thr leu glu glu trp gln ala glu his glu asp

3061/999 ATT ATG AAG GCC ATT AAT TCT ACA TCC ACA GTA TCT GAC CCT TTC GCC AGC AAA GTG AAT
ile met lys ala ile asn ser thr ser thr val ser asp pro phe ala ser lys val asn

3121/1019 ACA TGC TGG GCT AAA GCT ATT ATA CCC ATC CTA AGA ACG GCA GGG ATA GAA CTT ACA TTC
thr cys trp ala lys ala ile ile pro ile leu arg thr ala gly ile glu leu thr phe

3181/1039 GAG CAG TGG GAA GAT CTA TTC CCG CAA TTT CGT AAT GAC CAA CCT TAC TCC GTG ATG TAT
glu gln trp glu asp leu phe pro gln phe arg asn asp gln pro tyr ser val met tyr

3241/1059 GCC CTA GAT GTG ATA TGT ACC AAG ATG TTC GGC ATG GAT CTG AGC AGT GGG ATC TTC TCT
ala leu asp val ile cys thr lys met phe gly met asp leu ser ser gly ile phe ser

3301/1079 CGT CCT GAG ATA CCT CTA ACG TTC CAT CCC GCG GAC GTC GGC CGA GTG AGA GCT CAC TGG
arg pro glu ile pro leu thr phe his pro ala asp val gly arg val arg ala his trp

Figure 7c. See legend on last page of this sequence

```
3361/1099  GAT AAC TCC CCA GGA GGG CAG AAG TTT GGG TAT AAC AAG GCG GTA ATC CCA ACT TGC AAG
           asp asn ser pro gly gly gln lys phe gly tyr asn lys ala val ile pro thr cys lys

3421/1119  AAA TAC CCA GTG TAC TTA AGA GCA GGA AAA GGG GAC CAA ATA CTC CCC ATA TAT GGC AGA
           lys tyr pro val tyr leu arg ala gly lys gly asp gln ile leu pro ile tyr gly arg

3481/1139  GTT TCA GTC CCA TCG GCA CGG AAC AAT TTA GTT CCC TTA AAC AGA AAT CTA CCA CAC TCG
           val ser val pro ser ala arg asn asn leu val pro leu asn arg asn leu pro his ser

3541/1159  CTA ACT GCA AGC CTG CAG AAA AAA GAA GCA GCT CCC TTG CAC AAG TTC CTT AAC CAA CTA
           leu thr ala ser leu gln lys lys glu ala ala pro leu his lys phe leu asn gln leu

3601/1179  CCA GGA CAC AGT ATG CTG CTG GTC TCT AAG GAA ACA TGC TAT TGC GTG TCC AAG CGA ATC
           pro gly his ser met leu leu val ser lys glu thr cys tyr cys val ser lys arg ile

3661/1199  ACA TGG GTC GCT CCG CTG GGA GTC AGA GGA GCT GAC CAC AAC CAT GAC CTG CAT TTC GGG
           thr trp val ala pro leu gly val arg gly ala asp his asn his asp leu his phe gly

3721/1219  TTC CCA CCA CTG TCC AGA TAC GAC CTT GTG GTG GTT AAT ATG GGA CAA CCG TAC AGG TTC
           phe pro pro leu ser arg tyr asp leu val val val asn met gly gln pro tyr arg phe

3781/1239  CAT CAC TAC CAG CAG TGC GAG GAG CAT GCC GGC CTC ATG AGG ACG TTG GCC CGG TCA GCA
           his his tyr gln gln cys glu glu his ala gly leu met arg thr leu ala arg ser ala

3841/1259  CTC AAC TGC CTA AAA CCA GGA GGA ACA TTA GCC CTG AAA GCA TAT GGT TTC GCC GAC TCC
           leu asn cys leu lys pro gly gly thr leu ala leu lys ala tyr gly phe ala asp ser

3901/1279  AAT AGT GAG GAC GTT GTT CTG TCT TTA GCG AGG AAA TTC GTG CGG GCA TCC GCA GTG AGA
           asn ser glu asp val val leu ser leu ala arg lys phe val arg ala ser ala val arg

3961/1299  CCA TCG TGT ACA CAG TTT AAC ACA GAG ATG TTC TTT GTA TTT AGG CAG CTG GAC AAC GAT
           pro ser cys thr gln phe asn thr glu met phe phe val phe arg gln leu asp asn asp

4021/1319  CGT GAG CGC CAA TTC ACT CAG CAT CAC TTG AAT TTA GCA GTA TCC AAT ATA TTC GAC AAT
           arg glu arg gln phe thr gln his his leu asn leu ala val ser asn ile phe asp asn

4081/1339  TAT AAA GAC GGA TCC GGA GCA GCT CCT TCT TAT CGC GTT AAG AGA ATG AAT ATC GCA GAC
           tyr lys asp gly ser gly ala ala pro ser tyr arg val lys arg met asn ile ala asp

4141/1359  TGC ACA GAA GAA GCA GTG GTG AAC GCA GCT AAC GCG CGG GGA AAA CCT GGG GAC GGA GTA
           cys thr glu glu ala val val asn ala ala asn ala arg gly lys pro gly asp gly val

4201/1379  TGC AGA GCT ATC TTC AAA AAG TGG CCG AAG TCA TTT GAG AAC GCT ACC ACT GAA GTG GAA
           cys arg ala ile phe lys lys trp pro lys ser phe glu asn ala thr thr glu val glu

4261/1399  ACC GCG GTC ATG AAA CCA TGC CAC AAC AAG GTT GTT ATA CAT GCA GTG GGT CCT GAT TTT
           thr ala val met lys pro cys his asn lys val val ile his ala val gly pro asp phe

4321/1419  AGA AAG TAC ACG TTG GAG GAA GCG ACG AAG CTA CTG CAG AAC GCA TAC CAT GAT GTG GCA
           arg lys tyr thr leu glu glu ala thr lys leu leu gln asn ala tyr his asp val ala

4381/1439  AAG ATA GTG AAC GAG AAA GGC ATC TCC TCG GTA GCT ATA CCG CTG CTC TCA ACA GGT ATC
           lys ile val asn glu lys gly ile ser ser val ala ile pro leu leu ser thr gly ile

4441/1459  TAT GCT GCC GGA GCT GAT CGC CTG GAT CTC TCG CTG AGA TGT CTT TTC ACC GCG CTG GAT
           tyr ala ala gly ala asp arg leu asp leu ser leu arg cys leu phe thr ala leu asp
```

Figure 7d. See legend on last page of this sequence

```
4501/1479  CGT ACG CAT GCG GAT GTC ACA ATA TAT TGC CTA GAT AAG AAG TGG GAG CAA CGC ATA GCA
           arg thr asp ala asp val thr ile tyr cys leu asp lys lys trp glu gln arg ile ala

4561/1499  GAT GCT ATT AGG ATG CGA GAA CAA GTA ACT GAA TTA AAA GAT CCG GAC ATA GAG ATA GAT
           asp ala ile arg met arg glu gln val thr glu leu lys asp pro asp ile glu ile asp

4621/1519  GAA GGA TTA ACC CGG GTA CAC CCA GAT AGC TGC CTC AAG GAT CAC ATA GGC TAC AGT ACC
           glu gly leu thr arg val his pro asp ser cys leu lys asp his ile gly tyr ser thr

4681/1539  CAG TAT GGG AAA TTG TAC TCA TAC TTT GAA GGT ACT AAA TTC CAC CAA ACC GCA AAA GAC
           gln tyr gly lys leu tyr ser tyr phe glu gly thr lys phe his gln thr ala lys asp

4741/1559  ATA GCC GAG ATT CGT GCG CTG TTT CCT GAT GTA CAA GCC GCT AAC GAA CAA ATC TGC CTG
           ile ala glu ile arg ala leu phe pro asp val gln ala ala asn glu gln ile cys leu

4801/1579  TAC ACT TTA GGC GAA CCG ATG GAG TCC ATA CGC GAA AAG TGC CCA GTC GAA GAC TCC CCG
           tyr thr leu gly glu pro met glu ser ile arg glu lys cys pro val glu asp ser pro

4861/1599  GCA TCA GCA CCT CCT AAG ACA ATA CCT TGC CTA TGT ATG TAT GCT ATG ACA GCC GAA CGT
           ala ser ala pro pro lys thr ile pro cys leu cys met tyr ala met thr ala glu arg

4921/1619  ATT TGC CGC GTA CGC AGT AAC TCC GTA ACG AAC ATA ACG GTG TGC TCA TCC TTT CCG TTA
           ile cys arg val arg ser asn ser val thr asn ile thr val cys ser ser phe pro leu

4981/1639  CCC AAG TAC CGA ATA AAG AAC GTA CAA AAG ATA CAG TGC ACG AAA GTG
           pro lys tyr arg ile lys asn val gln lys ile gln cys thr lys val
```

Figure 7e. Translated sequence of Aura virus. This sequence starts near the 5'terminus of the genome, although the exact 5' end is not known. The translated sequence shown encompasses nsP1, nsP2, and the N-terminal (conserved) region of nsP3. Nucleotides are numbered from the beginning of the sequence; amino acids are numbered from the beginning of the open reading frame.

SEQUENCE ANALYSIS OF WHATAROA VIRUS.

We have obtained most of the sequence of Whataroa virus RNA, 11.7 kb in length. This sequence is being assembled to give the complete sequence of this virus RNA. We were interested in this virus because it represents a geographically isolated Sindbis-like virus, being found in New Zealand and presumably transferred there by migratory birds.

The sequencesof a stretch of the nonstructural protein coding region of the Whataroa genome is shown in Figs 8. The sequence begins near the beginning of the nsP2 gene and continues through to the end of the nsP2 region of the virus genome, a stretch of about 2000 nucleotides. From the analysis of this sequence, Whataroa virus can clearly be considered to be a strain of Sindbis virus that has spread to New Zealand. The amino acid sequence deduced from the nucleotide sequence in Fig. 8 is compared to that of the AR339 strain of Sindbis virus, isolated from Egypt in 1952, in Fig. 9. These amino acid sequences are 84% identical. Furthermore, we found that Whataroa virus RNA has the characteristic 3' NTR of the Sindbis viruses.

SEQUENCE ANALYSIS OF OTHER ALPHAVIRUSES

We have obtained the nucleotide sequence encoding the nsP3 and nsP4 genes of several other alphaviruses, in order to examine the relationships of viruses isolated from Australia, India, and South Africa to other alphaviruses. Sequences of this region for Sindbis virus isolated from India in 1953 is shown in Fig. 10, that for a Sindbis virus isolated in Australia in 1975 is shown in Fig. 11, and that for a Sindbis virus isolated from South Africa in 1963 is shown in Fig. 12. The South African isolated came from a human patient exhibiting symptoms of

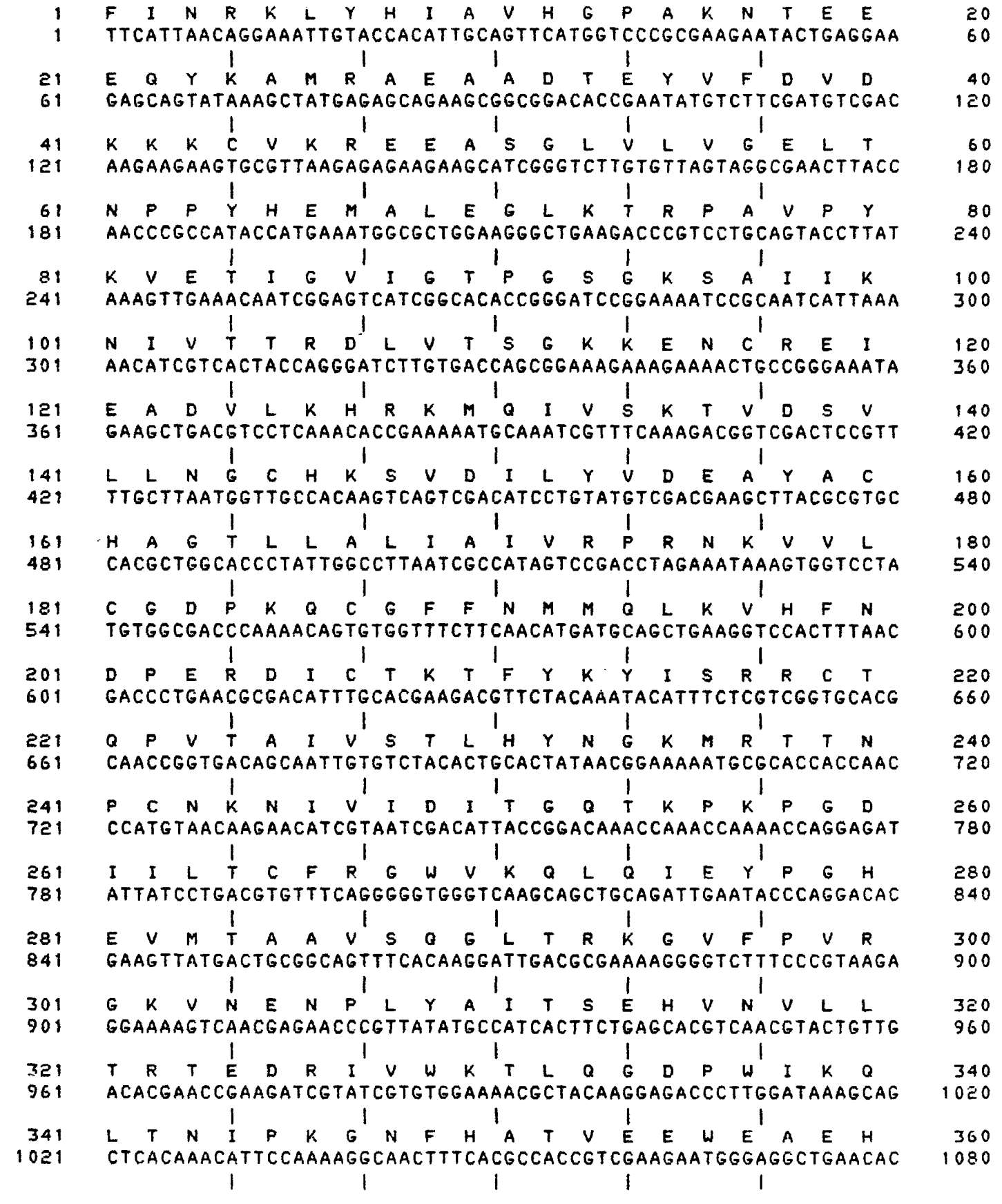


Figure 8a. See legend on next page.

```
 361   K  G  I  M  E  A  I  I  S  F  A  F  K  S  N  F  F  S  L  K      380
1081   AAGGGAATAATGGAGGCTATCACTAGCCCGGCCCCCCGCAGCAACCCTTTCAGCTGTAAG   1140
         |     |     |     |     |     |     |     |     |     |
 381   T  N  V  C  W  A  K  A  L  E  P  I  L  S  T  A  G  I  S  L      400
1141   ACAAACGTGTGCTGGGCGAAGGCACTAGAACCTATACTATCGACCGCTGGCATATCACTA   1200
         |     |     |     |     |     |     |     |     |     |
 401   T  G  C  Q  W  A  D  L  F  P  Q  F  E  D  D  K  P  H  S  A      420
1201   ACTGGATGTCAGTGGGCAGATTTGTTTCCGCAATTTGAAGATGACAAACCACATTCGGCC   1260
         |     |     |     |     |     |     |     |     |     |
 421   I  Y  A  L  D  V  I  C  V  K  F  F  G  M  D  L  T  S  G  I      440
1261   ATATACGCTCTAGACGTCATTTGCGTAAAGTTCTTTGGCATGGATTTAACTAGCGGCATA   1320
         |     |     |     |     |     |     |     |     |     |
 441   F  S  K  P  L  I  P  L  T  Y  H  P  A  E  G  D  R  K  T  A      460
1321   TTTTCAAAACCGTTGATCCCATTGACTTATCACCCCGCCGAAGGGGACCGGAAGACAGCG   1380
         |     |     |     |     |     |     |     |     |     |
 461   H  W  D  N  S  P  G  Q  R  K  Y  G  F  D  K  A  V  V  A  E      480
1381   CACTGGGACAACAGTCCAGGCCAACGAAAGTACGGGTTTGACAAAGCCGTTGTAGCTGAA   1440
         |     |     |     |     |     |     |     |     |     |
 481   L  S  R  R  F  P  V  F  C  M  A  D  K  G  V  Q  L  D  L  Q      500
1441   TTGTCCCGCAGATTCCCAGTATTCTGCATGGCAGACAAAGGAGTGCAACTGGACCTACAG   1500
         |     |     |     |     |     |     |     |     |     |
 501   T  G  R  T  R  V  V  ?  S  R  F  N  L  V  P  F  N  R  N  L      520
1501   ACGGGCCGNACGCGCGTAGTCNCGTCACGCTTCAACCTTGTGCCATTTAACAGAAATCTG   1560
         |     |     |     |     |     |     |     |     |     |
 521   P  H  S  L  V  P  E  Y  K  T  Q  T  P  G  Q  L  S  A  F  I      540
1561   CCCCACTCGCTTGTCCCGGAGTATAAAACACAAACTCCAGGTCAGCTAAGCGCCTTTATC   1620
         |     |     |     |     |     |     |     |     |     |
 541   R  Q  F  K  Q  N  T  I  L  L  V  S  E  T  P  A  E  H  S  T      560
1621   CGCCAGTTTAAACAAAACACCATCCTGCTTGTATCTGAAACACCTGCCGAACATTCCACC   1680
         |     |     |     |     |     |     |     |     |     |
 561   K  S  V  E  W  I  A  P  L  G  T  L  G  A  T  K  C  Y  N  L      580
1681   AAATCTGTGGAATGGATTGCACCGCTGGGTACGCTTGGAGCCACCAAATGCTATAATTTA   1740
         |     |     |     |     |     |     |     |     |     |
 581   A  F  G  F  P  P  Q  S  R  Y  D  L  V  I  I  N  I  G  T  K      600
1741   GCATTCGGCTTTCCGCCTCAGTCGAGGTACGACCTAGTGATCATAAATATCGGTACAAAA   1800
         |     |     |     |     |     |     |     |     |     |
 601   F  R  H  H  H  Y  Q  Q  C  E  D  H  A  A  T  M  K  T  L  S      620
1801   TTCAGACACCACCACTATCAACAGTGCGAAGACCACGCCGCCACCATGAAGACACTGTCA   1860
         |     |     |     |     |     |     |     |     |     |
 621   R  S  A  L  N  C  L  N  P  G  G  T  L  V  V  K  A  Y  G  Y      640
1861   CGTTCCGCCCTTAATTGCCTGAACCCGGGTGGCACATTGGTGGTAAAAGCATATGGCTAC   1920
         |     |     |     |     |     |     |     |     |     |
 641   A  D  R  N  S  E  D  I  I  T  A  L  A  R  K  F  V  R  V  S      660
1921   GCGGACAGAAACAGTGAAGACATCATTACAGCCCTGGCACGAAAGTTCGTCAGGGTGTCC   1980
         |     |     |     |     |     |     |     |     |     |
 661   A  A  R  P  Q  C  V  S  S  N  T  E  M  Y  F  I  F  R  Q  L      680
1981   GCGGCCCGCCCACAGTGCGTCTCAAGCAATACAGAGATGTACTTCATTTTCAGACAACTG   2040
         |     |     |     |     |     |     |     |     |     |
 681   D  N  S  R  T  R  Q  F  T  P  H  H  L  N  C  V  V  S  S  V      700
2041   GACAACAGCAGAACACGTCAATTCACACCTCATCACCTCAACTGCGTCGTTTCGTCAGTG   2100
         |     |     |     |     |     |     |     |     |     |
 701   Y  E  G  T  R  D  G  V  G  A                                    710
2101   TACGAGGGAACAAGAGACGGAGTTGGTGCT                                  2130
         |     |
```

Figure 8b. Translated nucleotide sequence of Whataroa virus in the region encoding nonstructural protein nsP2. By homology with Sindbis virus, the sequence shown begins at amino acid 97 of nsP2 and continues to the nsP2/nsP3 cleavage site.

```
FINRKLYHIAVHGPAKNTEEEQYKAMRAEAADTEYVFDVDKKKCVKREEA
.V.......M.............VTK..L.E..........R...K...
         *          *          *          *          *
SGLVLVGELTNPPYHEMALEGLKTRPAVPYKVETIGVIGTPGSGKSAIIK
.....S.........L.....................................
         *          *          *          *          *
NIVTTRDLVTSGKKENCREIEADVLKHRKMQIVSKTVDSVLLNGCHKSVD
ST..A.....................RL.G...T.......M......A.E
         *          *          *          *          *
ILYVDEAYACHAGTLLALIAIVRPRNKVVLCGDPKQCGFFNMMQLKVHFN
V......F.....A...........K........M................
         *          *          *          *          *
DPERDICTKTFYKYISRRCTQPVTAIVSTLHYNGKMRTTNPCNKNIVIDI
H..K.............................D...K.....K...E....
         *          *          *          *          *
TGQTKPKPGDIILTCFRGWVKQLQIEYPGHEVMTAAVSQGLTRKGVFPVR
..A.......................D..........A.........YA...
         *          *          *          *          *
GKVNENPLYAITSEHVNVLLTRTEDRIVWKTLQGDPWIKQLTNIPKGNFH
Q.........................L.............P........Q
         *          *          *          *          *
ATVEEWEAEHKGIMEAITSPAPRSNPFSCKTNVCWAKALEPILSTAGISL
..I.D........IA..N..T..A....................A....V.
         *          *          *          *          *
TGCQWADLFPQFEDDKPHSAIYALDVICVKFFGMDLTSGIFSKPLIPLTY
.....SE.....A...............I..........L...QS.....
         *          *          *          *          *
HPAEGDRKTAHWDNSPGQRKYGFDKAVVAELSRRFPVFCMADKGVQLDLQ
...DSA.PV.........T....Y.H.IA..........QL.G..T.....
         *          *          *          *          *
TGRTRVV?SRFNLVPFNRNLPHSLVPEYKTQTPGQLSAFIRQFKQNTILL
......ISAQH....V......A......EKQ..PVKK.LN...HHSV.V
         *          *          *          *          *
VSETPAEHSTKSVEWIAPLGTLGATKCYNLAFGFPPQSRYDLVIINIGTK
...EKI.APR.RI.....I.IA..D.N..........A.....F......
         *          *          *          *          *
FRHHHYQQCEDHAATMKTLSRSALNCLNPGGTLVVKAYGYADRNSEDIIT
Y.N..F.........L.....................S..........VV.
         *          *          *          *          *
ALARKFVRVSAARPQCVSSNTEMYFIFRQLDNSRTRQFTPHHLNCVVSSV
..............D.........L.....................I....
         *          *          *          *          *
YEGTRDGVGA
..........
```

Figure 9. Aligned deduced amino acid sequences of the nonstructural protein regions of Whataroa virus and Sindbis virus, beginning with amino acid 97 of Sindbis virus nsP2. The upper sequence in each case is Whataroa virus, and amino acid identity in the Sindbis sequence is indicated with a dot.

```
   1 GCUCCGGCCUAUCGCUCGAAACGUGAGAACAUCGCCGAGUGCCUCGAAGAGGCCGUAGUU   60
     A  P  A  Y  R  S  K  R  E  N  I  A  E  C  L  E  E  A  V  V

  61 AAUGCCGCGAAUGCACUCGGACGGCCGGGCGAAGGGGUAUGCAAAGCCAUAUAUAAAAAA  120
     N  A  A  N  A  L  G  R  P  G  E  G  V  C  K  A  I  Y  K  K

 121 UGGCCUAAUAGUUUCGUCGAUUCCGCGACAGAGACUGGAACGGCUAAGCUAGUGUGCUGU  180
     W  P  N  S  F  V  D  S  A  T  E  T  G  T  A  K  L  V  C  C

 181 CAAGGAAAGAAAAUUAUCCACGCCGUCGGACCCGACUUCCGCAAACACUCCGAGGCAGAA  240
     Q  G  K  K  I  I  H  A  V  G  P  D  F  R  K  H  S  E  A  E

 241 GCACUGAAGAUUCUCCAGAACACAUACCACGCCAUAGCAGAUUUGGUUAACAAACAUGGA  300
     A  L  K  I  L  Q  N  T  Y  H  A  I  A  D  L  V  N  K  H  G

 301 AUCAAGACUGUAGCGAUCCCGCUACUAUCCACCGGGAUUUACGCAGCGGGAAAAGACAGA  360
     I  K  T  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361 CUCGAGGUCUCCUUAAACUGUCUUACCACCGCCCUGGACAGAACAGACGCAGACGUCACA  420
     L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

 421 AUCUACUGUCUAGACAAAAAAUGGAAAGAAAGGAUCGAUGCGGUUAUACAAUUGAAGGAG  480
     I  Y  C  L  D  K  K  W  K  E  R  I  D  A  V  I  Q  L  K  E

 481 UCGGUGACGGAACUGAAGGAUGAGGAUAUGGAGAUCGACGAUGAGUUAGUAUGGAUCCAC  540
     S  V  T  E  L  K  D  E  D  M  E  I  D  D  E  L  V  W  I  H

 541 CCGGAUAGUUGUCUCAAGGGCAGGAAAGGGUAUAGCACAACAAAAGGUAAACUUUAUUCG  600
     P  D  S  C  L  K  G  R  K  G  Y  S  T  T  K  G  K  L  Y  S

 601 UACUUUGAGGGGACUAAGUUUCAUCAGGCAGCAAAAGACAUGGCGGAGAUUAAAGUACUU  660
     Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

 661 UUUCCCGAUGAGCAAGAGUGCAACGAGCAGUUGUGUGCAUACAUCCUUGGUGAAACCAUG  720
     F  P  D  E  Q  E  C  N  E  Q  L  C  A  Y  I  L  G  E  T  M

 721 GAAGCCAUCAGGGAAAAAUGUCCAGUGGACUUUAAUCCGUCGUCCAGUCCGCCGAAGACA  780
     E  A  I  R  E  K  C  P  V  D  F  N  P  S  S  S  P  P  K  T

 781 CUCCCCUGUUUGUGCAUGUAUGCCAUGACGCCUGAGAGAGUGCACCGUCUGCGUAGCAAC  840
     L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841 AACGUCAAGUCCAUCACAGUGUGUUCGUCUACCCCACUUCCGAAGCACAAGAUCAAGAAC  900
     N  V  K  S  I  T  V  C  S  S  T  P  L  P  K  H  K  I  K  N

 901 GUUCAGAAAGUACAGUGCACGAAAGUGGUCUUGUUCAAUCCACAGACCCCUGAAUUUGUC  960
     V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  Q  T  P  E  F  V

 961 CCUGCCCGUAAGUACAUAGAAGCACAACCAAAAGACGUAAGCCAAGAUGCAGAAGAAAGC 1020
     P  A  R  K  Y  I  E  A  Q  P  K  D  V  S  Q  D  A  E  E  S

1021 CCUGCCGCAGCCGCCCGAGAUAACACCUCACGGGACGUAACAGACAUAUCCCUGGAUGUG 1080
     P  A  A  A  A  R  D  N  T  S  R  D  V  T  D  I  S  L  D  V

1081 GAAGAAAGUCAAGCCGCAGCCGGCCAACCAGAGGAGCGCUCGGGGGACAACACUUCCCGG 1140
     E  E  S  Q  A  A  A  G  Q  P  E  E  R  S  G  D  N  T  S  R

1141 GAUGUAACAGAUAUAUCCCUAGAUCACGACAGCGAUAGUGAGGUGGGCUCCAUCUUCUCU 1200
     D  V  T  D  I  S  L  D  H  D  S  D  S  E  V  G  S  I  F  S

1201 AACCUUAGCUGCUCCAGUCAAUCCAUCACUAGUAUGGACAGCUGGUCCUCCGGACCGGGA 1260
     N  L  S  C  S  S  Q  S  I  T  S  M  D  S  W  S  S  G  P  G
```

Figure 10a. See legend at the end of this sequence.

```
1261  UCGAUCACGAUAAACGAGAACCGCACCAUUCAGGUCACGGCGGAGAUACACAAUGCUCCU  1320
      S  I  T  I  N  E  N  R  T  I  Q  V  T  A  E  I  H  N  A  P

1321  GCCGCGUUGCCUGUUCCACCACCACGCCUUAAGAAACUGGCACGCUUAGCAGCCCAGAAG  1380
      A  A  L  P  V  P  P  P  R  L  K  K  L  A  R  L  A  A  Q  K

1381  CCCAAUCCGCCAUCCGACCCGCCUUCGACGGUCGAGGACGUGUCGAUGCGCUUGUCCUUC  1440
      P  N  P  P  S  D  P  P  S  T  V  E  D  V  S  M  R  L  S  F

1441  CCUGCCACGGUGUCGUUCGGAUCAUUCUCCGACGGAGAAGUCGACGACCUUAGCCGCGAU  1500
      P  A  T  V  S  F  G  S  F  S  D  G  E  V  D  D  L  S  R  D

1501  AAAGCAGUGUCAGAACCGGUGGUCUUUGGUGCUUUCGAGCCUGGAGAGGUAACCUCUAUC  1560
      K  A  V  S  E  P  V  V  F  G  A  F  E  P  G  E  V  T  S  I

1561  AUCGAAUCAAGGUCUGUCGUGUCAUUCCCCGUGCAUAAACGCCGGCGCAGAAGACGGGGC  1620
      I  E  S  R  S  V  V  S  F  P  V  H  K  R  R  R  R  R  R  G

1621  AAAAGAACCGAAUAUUGACUAACCGGGGUAGGUGGGUACAUCUUCUCAACUGACACGGGA  1680
      K  R  T  E  Y  *  L  T  G  V  G  G  Y  I  F  S  T  D  T  G

1681  CCGGGCCACCUCCAGAAGAAGUCAGUUCUGCAAAACCAGCUUACUGAACCGACCCUCGAG  1740
      P  G  H  L  Q  K  K  S  V  L  Q  N  Q  L  T  E  P  T  L  E

1741  CGCAAUCAAUUAGAACGAAUGUAUGCGCCCAGUCUCGAUGUCAAGAAAGAGGAACUUCUG  1800
      R  N  Q  L  E  R  M  Y  A  P  S  L  D  V  K  K  E  E  L  L

1801  AAACUUAAGUACCAAAUGAUGCCCACCGAAGCCAAUAAAAGUAGGUACCAGUCUAGAAAG  1860
      K  L  K  Y  Q  M  M  P  T  E  A  N  K  S  R  Y  Q  S  R  K

1861  GUUGAAAAUCAAAAAGCGGUAACCACCGAGAGGUUACUGUCGGGACUGAAGAUGUACAUC  1920
      V  E  N  Q  K  A  V  T  T  E  R  L  L  S  G  L  K  M  Y  I

1921  CACUCAGAGAACCAACCUGAGUGUUAUAAGGUCACUUAUCCGAAACCGUCGUACUCCAGC  1980
      H  S  E  N  Q  P  E  C  Y  K  V  T  Y  P  K  P  S  Y  S  S

1981  AGUGUCCCUCUUAGUUACCAGAACCCUGAAUUCGCCGUAGCUGUUUGCAAUAACUACCUG  2040
      S  V  P  L  S  Y  Q  N  P  E  F  A  V  A  V  C  N  N  Y  L

2041  CAUGAGAACUACCCGACGGUUGCCUCCUAUCAGAUUACGGACGAAUAUGAUGCCUACCUC  2100
      H  E  N  Y  P  T  V  A  S  Y  Q  I  T  D  E  Y  D  A  Y  L

2101  GACAUGGUGGACGGCACUGUUGCGUGUCUCGACACUGCAACAUUCUGCCCUGCGAAAUUA  2160
      D  M  V  D  G  T  V  A  C  L  D  T  A  T  F  C  P  A  K  L

2161  CGUAGCUUUCCGAAGAAACAUGAGUACCGCGCACCUAACAUCAGGAGUGCCGUGCCGUCU  2220
      R  S  F  P  K  K  H  E  Y  R  A  P  N  I  R  S  A  V  P  S

2221  GCUAUGCAGAACACUCUACAGAACGUCCUGAAUGCAGCAACAAAGAGGAAUUGCAACGUU  2280
      A  M  Q  N  T  L  Q  N  V  L  N  A  A  T  K  R  N  C  N  V

2281  ACUCAGAUGAGAGAACUACCGACCCUAGACUCCGCGACCUUUAACGUGGAAUGCUUCCGA  2340
      T  Q  M  R  E  L  P  T  L  D  S  A  T  F  N  V  E  C  F  R

2341  AAGUACGCGUGCAAUGACGAGUAUUGGGCUGAAUUCUCCGAAAAACCAAUCAGGAUCACC  2400
      K  Y  A  C  N  D  E  Y  W  A  E  F  S  E  K  P  I  R  I  T

2401  ACGGAGUUUGUUACGGCGUACGUGGCGAGAUUGAAGGGACCAAAGGCUGCUGCUCUGUUU  2460
      T  E  F  V  T  A  Y  V  A  R  L  K  G  P  K  A  A  A  L  F

2461  GCAAAAACGCAUAACCUAGUCCCAUUGCAAGAAGUACCUAUGGACAGGUUUGUGAUGGAC  2520
      A  K  T  H  N  L  V  P  L  Q  E  V  P  M  D  R  F  V  M  D
```

Figure 10b. See legend at the end of this sequence.

```
2521  AUGAAGCGAGAUGUCAAGGUGACUCCGGGCACAAAACACACCGAAGAAAGGCCUAAGGUG  2580
      M  K  R  D  V  K  V  T  P  G  T  K  H  T  E  E  R  P  K  V

2581  CAGGUAAUCCAAGCGGCUGAGCCUUUUGCUACAGCCUACCUUUGUGGCAUCCACCGAGAG  2640
      Q  V  I  Q  A  A  E  P  F  A  T  A  Y  L  C  G  I  H  R  E

2641  CUGGUACGCCGGCUUACCGCGGUUCUACUCCCGAACGUACACACCCUGUUUGACAUGUCU  2700
      L  V  R  R  L  T  A  V  L  L  P  N  V  H  T  L  F  D  M  S

2701  GCGGAGGAUUUCGACGCGAUUAUUGCCGAGCAUUUCCGACAAGGUGACGCCGUGCUCGAG  2760
      A  E  D  F  D  A  I  I  A  E  H  F  R  Q  G  D  A  V  L  E

2761  ACAGACAUCGCGUCAUUCGAUAAGAGUCAGGACGAUGCGAUGGCCCUGACUGGGCUGAUG  2820
      T  D  I  A  S  F  D  K  S  Q  D  D  A  M  A  L  T  G  L  M

2821  AUCCUGGAGGACCUCGGCGUCGAUCAACCGCUGCUGGACCUCAUCGAGUGUGCCUUCGGA  2880
      I  L  E  D  L  G  V  D  Q  P  L  L  D  L  I  E  C  A  F  G

2881  GAAAUAUCAUCUACGCAUCUGCCUACUGGGACACGGUUUAAGUUCGGCUCAAUGAUGAAA  2940
      E  I  S  S  T  H  L  P  T  G  T  R  F  K  F  G  S  M  M  K

2941  UCCGGAAUGUUUCUUACGCUCUUCGUGAACACCAUCUUGAAUGUCGUGAUCGCUAGUCGC  3000
      S  G  M  F  L  T  L  F  V  N  T  I  L  N  V  V  I  A  S  R

3001  GUGCUUGAGCACAGGUUAACAGGAUCACGAUGUGCCGCAUUCAUUGGAGACGAUAACAUC  3060
      V  L  E  H  R  L  T  G  S  R  C  A  A  F  I  G  D  D  N  I

3061  AUCCACGGCGUGGUAUCAGACAAGGAAAUGGCCGAAAGGUGCGCCACUUGGCUGAAUAUG  3120
      I  H  G  V  V  S  D  K  E  M  A  E  R  C  A  T  W  L  N  M

3121  GAGGUAAAAAUCAUUGACGCGGUGAUCGGCGAGCGUCCUCCGUAUUUCUGUGGUGGCUUU  3180
      E  V  K  I  I  D  A  V  I  G  E  R  P  P  Y  F  C  G  G  F

3181  AUACUACAGGACUCUGUCACCCAAACAGCCUGUCGAGUGGCUGACCCCCUAAAAAGACUG  3240
      I  L  Q  D  S  V  T  Q  T  A  C  R  V  A  D  P  L  K  R  L

3241  UUCAAGCUAGGAAAACCUUUGCCCGCAGAUGAUGACCAAGAUGAAGACAGAAGAAGGGCU  3300
      F  K  L  G  K  P  L  P  A  D  D  D  Q  D  E  D  R  R  R  A

3301  UUGCUGGAUGAGACUAAGGCGUGGUUUAGAGUGGGCAUAACCGAAACAUUGGCUACUGCG  3360
      L  L  D  E  T  K  A  W  F  R  V  G  I  T  E  T  L  A  T  A

3361  GUAGCAACGCGGUACGAAGUUGAUAACAUCACGCCUGUCCUGCUGGCACUGAGGACCCUU  3420
      V  A  T  R  Y  E  V  D  N  I  T  P  V  L  L  A  L  R  T  L

3421  GCGCAAAGCAAGAGAUCCUUUCAGUCCAUAAGAGGGGAAAUGAAGCAUCUCUACGGUGGU  3480
      A  Q  S  K  R  S  F  Q  S  I  R  G  E  M  K  H  L  Y  G  G

3481  CCUAAAUAG  3489
      P  K  *
```

Figure 10c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 of Sindbis A1036, isolated in India in 1953. The sequence has been translated using the single letter amino acid code.

```
1     GCUCCGGCCUACCGCUCGAAACGUGAGAAUAUCGCCGAAUGCCUUGAAGAGGCCGUAGUU   60
      A  P  A  Y  R  S  K  R  E  N  I  A  E  C  L  E  E  A  V  V

61    AACGCCGCGAACCCACUCGGACGUCCGGGCGAAGGGGUGUGUAAAGCCAUAUAUAAAAAA  120
      N  A  A  N  P  L  G  R  P  G  E  G  V  C  K  A  I  Y  K  K

121   UGGCCCAAUAGUUUUGUCGAUUCUGCGACAGAGACUGGAACAGCUAAGCUAGUGUGCUGU  180
      W  P  N  S  F  V  D  S  A  T  E  T  G  T  A  K  L  V  C  C

181   CAAGGAAAAAAGAUUAUCCAUGCCGUCGGACCUGACUUCCGUAAACACCCCGAGGCAGAA  240
      Q  G  K  K  I  I  H  A  V  G  P  D  F  R  K  H  P  E  A  E

241   GCGCUGAAGAUUCUCCAGAACACAUACCACGCCAUCGCAGAUUUGGUUAACAAACAUGGA  300
      A  L  K  I  L  Q  N  T  Y  H  A  I  A  D  L  V  N  K  H  G

301   AUCAAGACCGUAGCGAUCCCGCUUCUAUCCACCGGGAUUUACGCAGCGGGAAAAGACAGA  360
      I  K  T  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

361   CUUGAGGUCUCUUUAAACUGCCUCACUACCGCCCUGGACAGAACUGACGCAGACGUCACA  420
      L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

421   AUCUACUGCCUUGACAAAAAAUGGAAAGAACGGAUUGAUGCGUUUAUACAGUUGAAGGAG  480
      I  Y  C  L  D  K  K  W  K  E  R  I  D  A  F  I  Q  L  K  E

481   UCGGUGACGGAACUGAAGGAUGAUGACAUGGAGAUCGACGACGAAUUAGUAUGGAUCCAC  540
      S  V  T  E  L  K  D  D  D  M  E  I  D  D  E  L  V  W  I  H

541   CCGGAUAGUUGCCUCAAGGGUAGGAAAGGGUUUAGUACGACGAAGGGCAAGCUCUACUCG  600
      P  D  S  C  L  K  G  R  K  G  F  S  T  T  K  G  K  L  Y  S

601   UACUUUGAGGGGACUAAAUUUCAUCAAGCAGCAAAAGACAUGGCUGAGAUCAAGGUACUU  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

661   UUUCCCGAUGAGCAAGAGUGCAACGAGCAACUGUGUGCAUACAUUCUAGGCGAAACCAUG  720
      F  P  D  E  Q  E  C  N  E  Q  L  C  A  Y  I  L  G  E  T  M

721   GAAGCCAUCAGGGAAAAAUGUCCAGUGGACUUUAAUCCGUCGUCCAGUCCGCCGAAGACG  780
      E  A  I  R  E  K  C  P  V  D  F  N  P  S  S  S  P  P  K  T

781   CUUCCCUGUUUGUGUAUGUACGCCAUGACGCCCGAGAGAGUGCACCGCUUGCGUAGCAAU  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

841   AACGUCAAAUCCAUCACAGUAUGCUCGUCAACCCCGCUUCCGAAGCACAAAAUUAAGAAC  900
      N  V  K  S  I  T  V  C  S  S  T  P  L  P  K  H  K  I  K  N

901   GUUCAGAAAGUACAGUGCACGAAAGUAGUCCUAUUCAACCCACAAACGCCUGAAUUUGUC  960
      V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  Q  T  P  E  F  V

961   CCUGCCCGCAAGUACAUAGAAACACAACCGAAGGACGACAGUCAAGAGGCGGAAGAAAAC  1020
      P  A  R  K  Y  I  E  T  Q  P  K  D  D  S  Q  E  A  E  E  N

1021  CCUGCCGCAGCCGAUAACACUUCACGGGAUGUAACAGACGUAUCUCUAGAUGUGGAAGGA  1080
      P  A  A  A  D  N  T  S  R  D  V  T  D  V  S  L  D  V  E  G

1081  GAUCGCGUUGCGGCCAACCGAUCAGAGGUGCACUCAGAGGACAACACCUCCCGAGAUGUA  1140
      D  R  V  A  A  N  R  S  E  V  H  S  E  D  N  T  S  R  D  V

1141  ACAGACAUAUCUCUAGACCACAACAGUGAUAGCGAGGUGGGCUCCAUUUUCUCUGACCUC  1200
      T  D  I  S  L  D  H  N  S  D  S  E  V  G  S  I  F  S  D  L

1201  AGCUGCUCCAGUCAUUCCAUCACCAGCAUGGACAGCUGGUCCUCCGGACCGAGCUCGAUC  1260
      S  C  S  S  H  S  I  T  S  M  D  S  W  S  S  G  P  S  S  I
```

Figure 11a. See legend on the last page of this sequence.

```
1261  AUGCUAAACGGGAAUCACACCAUCCAGGUCACGGCAGAGAUACACAACGCUCCUGCUGCA  1320
      M  L  N  G  N  H  T  I  Q  V  T  A  E  I  H  N  A  P  A  A

1321  CCGCCCGUACCACCACCACGCCUCAAGAAACUGGCGCGCUUGGCAGCUCAGAAGUCCGAU  1380
      P  P  V  P  P  P  R  L  K  K  L  A  R  L  A  A  Q  K  S  D

1381  CCGCCAUCCAGCCCGCCCUCAACGGUUGAGGACGUGUCGAUGCGCCUGUCAUUCCCUGCC  1440
      P  P  S  S  P  P  S  T  V  E  D  V  S  M  R  L  S  F  P  A

1441  ACGGUGUCAUUCGGAUCUUUUUCUGACGGCGAAGUCGACGAUCUUAGUCGCGAAAAAGCA  1500
      T  V  S  F  G  S  F  S  D  G  E  V  D  D  L  S  R  E  K  A

1501  GUGUCAGAACCAGUGGUCUUUGGUGCUUUCGAGCCAGGAGAGGUAACAUCUAUCAUUGAA  1560
      V  S  E  P  V  V  F  G  A  F  E  P  G  E  V  T  S  I  I  E

1561  GCAAGGUCUGUCGUGUCAUUCCCCGUGAAUAAACGCCGGCGCAGGAGACGGGGCCAAAAG  1620
      A  R  S  V  V  S  F  P  V  N  K  R  R  R  R  R  R  G  Q  K

1621  AAAACCGAAUAUUGACUAACCGGGGUAGGUGGGUAUAUCUUCUCGACUGACACGGGACCG  1680
      K  T  E  Y  *  L  T  G  V  G  G  Y  I  F  S  T  D  T  G  P

1681  GGUCACCUCCAGAAAAAAAUCGGUUCUACAAAACCAGCUUACGGAACCGACCCUCGAGCGU  1740
      G  H  L  Q  K  K  S  V  L  Q  N  Q  L  T  E  P  T  L  E  R

1741  AAUCAAUUAGAACGAGUGUAUGCACCCAGUCUUGAUGCCAAGAAAGAGGAACUCUUGAAA  1800
      N  Q  L  E  R  V  Y  A  P  S  L  D  A  K  K  E  E  L  L  K

1801  CUCAAGUACCAAAUGAUGCCCACCGAAGCCAAUAAAAGUAGGUACCAGUCUAGAAAGGUA  1860
      L  K  Y  Q  M  M  P  T  E  A  N  K  S  R  Y  Q  S  R  K  V

1861  GAAAACCAAAAAGCCGUAACCACCGAGAGGUUACUGUCGGGAUUGAAGAUGUACAUUCAC  1920
      E  N  Q  K  A  V  T  T  E  R  L  L  S  G  L  K  M  Y  I  H

1921  UCAGAGAACCAACCCGAGUGUUACAAGGUCACCUAUCCGAAACCGUCGUACUCUAGCAGU  1980
      S  E  N  Q  P  E  C  Y  K  V  T  Y  P  K  P  S  Y  S  S  S

1981  GUUCCCCUUAGUUACCAGAGCCCCGAAUUCGCCGUAGCCGUCUGCAAUAACUACCUGCAU  2040
      V  P  L  S  Y  Q  S  P  E  F  A  V  A  V  C  N  N  Y  L  H

2041  GAGAAUUAUCCAACGGUUGCCUCCUAUCAGAUUACGGAUGAAUAUGACGCCUACCUUGAC  2100
      E  N  Y  P  T  V  A  S  Y  Q  I  T  D  E  Y  D  A  Y  L  D

2101  AUGGUGGACGGCACCGUAGCGUGUCUCGACACCGCUACAUUUUGCCCCGCGAAAUUACGC  2160
      M  V  D  G  T  V  A  C  L  D  T  A  T  F  C  P  A  K  L  R

2161  AGCUUCCCGAAGAAACACGAGUACCGAGAACCUAACAUCAGGAGCGCCGUACCGUCCGCU  2220
      S  F  P  K  K  H  E  Y  R  E  P  N  I  R  S  A  V  P  S  A

2221  AUGCAGAACACUCUACAGAACGUCCUGAACGCAGCAACAAAGAGGAAUUGCAAUGUUACU  2280
      M  Q  N  T  L  Q  N  V  L  N  A  A  T  K  R  N  C  N  V  T

2281  CAGAUGAGAGAACUACCGACUUUAGACUCCGCAACCUUUAAUGUGGAAUGCUUUCGAAAG  2340
      Q  M  R  E  L  P  T  L  D  S  A  T  F  N  V  E  C  F  R  K

2341  UACGCGUGCAACGACGAGUAUUGGGCUGAAUUCUCCGAAAAACCAAUUAGGAUCACCACA  2400
      Y  A  C  N  D  E  Y  W  A  E  F  S  E  K  P  I  R  I  T  T

2401  GAGUUUGUCACGGCGUACGUGGCGAGAUUGAAGGGACCAAAGGCUGCUGCACUGUUUGCU  2460
      E  F  V  T  A  Y  V  A  R  L  K  G  P  K  A  A  A  L  F  A

2461  AAAACGCAUAACCUAGUCCCACUGCAAGAAGUACCUAUGGACAGGUUUGUGAUGGACAUG  2520
      K  T  H  N  L  V  P  L  Q  E  V  P  M  D  R  F  V  M  D  M
```

Figure 11b. See legend on the last page of this sequence.

```
2521  AAGCGAGACGUUAAGGUGACUCCGGGCACGAAGCACACCGAAGAAAGACCCAAAGUGCAG  2580
      K  R  D  V  K  V  T  P  G  T  K  H  T  E  E  R  P  K  V  Q

2581  GUAAUCCAAGCGGCAGAGCCUCUAGCUACAGCCUAUUUAUGCGGCAUCCACCGUGAGCUG  2640
      V  I  Q  A  A  E  P  L  A  T  A  Y  L  C  G  I  H  R  E  L

2641  GUACGCAGGCUUACCGCAGUCCUGCUUCCGAACGUACACACCCUUUUUGAUAUGUCUGCG  2700
      V  R  R  L  T  A  V  L  L  P  N  V  H  T  L  F  D  M  S  A

2701  GAAGAUUUCGAUGCUAUCAUUGCCGAGCAUUUUCACCAGGGUGACGCUGUGCUCGAGACA  2760
      E  D  F  D  A  I  I  A  E  H  F  H  Q  G  D  A  V  L  E  T

2761  GACAUCGCGUCGUUCGAUAAGAGCCAAGACGAUGCGAUGGCCCUGACGGGGCUGAUGAUC  2820
      D  I  A  S  F  D  K  S  Q  D  D  A  M  A  L  T  G  L  M  I

2821  CUGGAGGACCUCGGAGUCGACCAGCCAUUGCUGGACCUCAUCGAGUGCGCCUUCGGGGAA  2880
      L  E  D  L  G  V  D  Q  P  L  L  D  L  I  E  C  A  F  G  E

2881  AUAUCAUCUACGCACCUGCCGACCGGGACACGGUUUAAGUUCGGCUCAAUGAUGAAAUCC  2940
      I  S  S  T  H  L  P  T  G  T  R  F  K  F  G  S  M  M  K  S

2941  GGAAUGUUCCUCACGCUCUUUGUGAACACCAUCUUGAAUGUCGUGAUAGCUAGUCGCGUG  3000
      G  M  F  L  T  L  F  V  N  T  I  L  N  V  V  I  A  S  R  V

3001  CUCGAGCACAGGUUAGCAGAAUCACGAUGCGCCGCAUUCAUCGGAGACGACAAUAUUAUU  3060
      L  E  H  R  L  A  E  S  R  C  A  A  F  I  G  D  D  N  I  I

3061  CACGGCGUGGUAUCCGAC[illegible]AAGAAAUGGCUGAAAGGUGCGCCACUUGGCUGAAUAUGGAG  3120
      H  G  V  V  S  [illegible]  K  E  M  A  E  R  C  A  T  W  L  N  M  E

3121  GUAAAAAUUA[illegible]CGCAGUAAUUGGCGAACGUCCUCCGUACUUCUGUGGCGGCUUUAUA  3180
      V  K  I  I  D  A  V  I  G  E  R  P  P  Y  F  C  G  G  F  I

3181  CUGCAGGACUCAGUCACCCAAACAGCCUGCCGAGUGGCGGACCCCCUAAAAAGAUUGUUC  3240
      L  Q  D  S  V  T  Q  T  A  C  R  V  A  D  P  L  K  R  L  F

3241  AAAUUAGGAAAACCAUUACCUGCAGAUGAUGACCAAGAUGAAGACAGAAGAAGGGCUCUG  3300
      K  L  G  K  P  L  P  A  D  D  D  Q  D  E  D  R  R  R  A  L

3301  CUGGAUGAGACCAAGGCGUGGUUUAGAGUGGGCAUAACUGAGACACUGGCUACUGCGGUA  3360
      L  D  E  T  K  A  W  F  R  V  G  I  T  E  T  L  A  T  A  V

3361  GCAACGCGGUAUGAAGUUGAUAACAUCACACCGGUCCUGCUGGCACUGAGGACCCUUGCG  3420
      A  T  R  Y  E  V  D  N  I  T  P  V  L  L  A  L  R  T  L  A

3421  CAAAGCAAGAGAUCUUUUCAGGCCAUAAGGGGGAAAAUGAAGCAUCUCUACGGUGGUCCU  3480
      Q  S  K  R  S  F  Q  A  I  R  G  K  M  K  H  L  Y  G  G  P

3481  AAAUAG  3486
      K  *
```

Figure 11c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 of an isolate of Sindbis virus isolated from a mosquito pool from Australia in 1975.

```
   1  GCACCGUCAUACCGCACUAAAAGGGAGAACAUUGCUGAUUGUCAAGAGGAAGCAGUUGUC   60
      A  P  S  Y  R  T  K  R  E  N  I  A  D  C  Q  E  E  A  V  V

  61  AAUGCAGCCAAUCCGCUGGGCAGACCAGGCGAAGGAGUCUGCCGUGCCAUCUAUAAACGU  120
      N  A  A  N  P  L  G  R  P  G  E  G  V  C  R  A  I  Y  K  R

 121  UGGCCGAACAGUUUCACCGAUUCAGCCACAGAGACCGGCACCGCAAAACUGACUGUGUGC  180
      W  P  N  S  F  T  D  S  A  T  E  T  G  T  A  K  L  T  V  C

 181  CAAGGAAAGAAAGUGAUCCACGCGGUUGGCCCUGAUUUCCGGAAACACCCAGAGGCAGAA  240
      Q  G  K  K  V  I  H  A  V  G  P  D  F  R  K  H  P  E  A  E

 241  GCCCUGAAAUUGCUGCAAAACGCCUACCAUGCAGUGGCAGACUUAGUAAAUGAACAUAAU  300
      A  L  K  L  L  Q  N  A  Y  H  A  V  A  D  L  V  N  E  H  N

 301  AUCAAGUCUGUCGCCAUCCCACUGCUAUCUACAGGCAUUUACGCAGCCGGAAAAGACCGC  360
      I  K  S  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361  CUUGAAGUAUCACUUAACUGCUUGACAACCGCGCUAGAUAGAACUGAUGCGGACGUAACC  420
      L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

 421  AUCUACUGCCUGGAUAAGAAGUGGAAGGAAAGAAUCGACGCGGUGCUCCAACUUAAGGAG  480
      I  Y  C  L  D  K  K  W  K  E  R  I  D  A  V  L  Q  L  K  E

 481  UCUGUAACAGAGCUGAAGGAUGAGGAUAUGGAGAUCGACGACGAGUUAGUAUGGAUCCAU  540
      S  V  T  E  L  K  D  E  D  M  E  I  D  D  E  L  V  W  I  H

 541  CCGGACAGUUGCCUGAAGGGAAGAAAGGGAUUCAGUACUACAAAAGGAAAGUUGUAUUCG  600
      P  D  S  C  L  K  G  R  K  G  F  S  T  T  K  G  K  L  Y  S

 601  UACUUUGAAGGCACCAAAUUCCAUCAAGCAGCAAAAGAUAUGGCGGAGAUAAAGGUCCUG  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

 661  UUCCCAAAUGACCAGGAAAGCAACGAGCAACUGUGUGCCUACAUAUUGGGGGAGACCAUG  720
      F  P  N  D  Q  E  S  N  E  Q  L  C  A  Y  I  L  G  E  T  M

 721  GAAGCAAUCCGCGAAAAAUGCCCGGUCGACCACAACCCGUCGUCUAGCCCGCCAAAAACG  780
      E  A  I  R  E  K  C  P  V  D  [illegible] N  P  S  S  S  P  P  K  T

 781  CUGCCGUGCCUCUGCAUGUAUGCCAUGA[illegible]CAGAAAGGGUCCACAGACUCAGAAGCAAC  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841  AACGUCAAAGAAGUUACAGUAUGCUCCUCCACCCCCCUUCCAAAGUACAAAAUCAAGAAC  900
      N  V  K  E  V  T  V  C  S  S  T  P  L  P  K  Y  K  I  K  N

 901  GUUCAGAAGGUUCAGUGCACAAAAGUAGUCCUGUUUAACCCGCAUACCCCUGCAUUCGUU  960
      V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  H  T  P  A  F  V

 961  CCCGCCCGUAAGUACAUAGAAGCGCCAGAACAGCCUGCAGCUCCGCCUGCACAGGCCGAG 1020
      P  A  R  K  Y  I  E  A  P  E  Q  P  A  A  P  P  A  Q  A  E

1021  GAGGCCCCCGAAGUUGCAGCAACACCAACACCACCUGCAGCUGAUAACACCUCGCUUGAU 1080
      E  A  P  E  V  A  A  T  P  T  P  P  A  A  D  N  T  S  L  D

1081  GUCACGGACAUCUCACUGGACAUGGAAGACAGUAGCGAAGGCUCACUCUUUUCGAGCUUU 1140
      V  T  D  I  S  L  D  M  E  D  S  S  E  G  S  L  F  S  S  F

1141  AGCGGAUCGGACAACUCUAUUACUAGUAUGGACAGUUGGUCGUCAGGACCUAGUUCACUA 1200
      S  G  S  D  N  S  I  T  S  M  D  S  W  S  S  G  P  S  S  L

1201  GAGAUAGUAGACCGAAGGCAGGUGGUGGUGGCUGACGUCCAUGCCGUCCAAGAGCCUGCC 1260
      E  I  V  D  R  R  Q  V  V  V  A  D  V  H  A  V  Q  E  P  A
```

Figure 12a. See legend on last page of this sequence.

```
1261  CCUGUUCCACCGCCAAGGCUAAAGAAGAUGGCCCGCCUGGCAGCGGCAAGAAUGCAGGAA  1320
      P  V  P  P  P  R  L  K  K  M  A  R  L  A  A  A  R  M  Q  E

1321  GAGCCAACUCCACCGGCAAGCACCAGCUCUGCGGACGAGUCCCUUCACCUUUCUUUUGGU  1380
      E  P  T  P  P  A  S  T  S  S  A  D  E  S  L  H  L  S  F  G

1381  GGGGUAUCCAUGUCCUUCGGAUCCCUUUUCGACGGAGAGAUGGCCCGCUUGGCAGCGGCA  1440
      G  V  S  M  S  F  G  S  L  F  D  G  E  M  A  R  L  A  A  A

1441  CAACCCCCGGCAAGUACAUGCCCUACGGAUGUGCCUAUGUCUUUCGGAUCGUUUUCCGAC  1500
      Q  P  P  A  S  T  C  P  T  D  V  P  M  S  F  G  S  F  S  D

1501  GGAGAGAUUGAGGAGCUGAGCCGCAGAGUAACCGAGUCUGAGCCCGUCCUGUUUGGGUCA  1560
      G  E  I  E  E  L  S  R  R  V  T  E  S  E  P  V  L  F  G  S

1561  UUUGAACCGGGCGAAGUGAACUCAAUUAUAUCGUCCCGAUCAGCCGUAUCUUUUCCACCA  1620
      F  E  P  G  E  V  N  S  I  I  S  S  R  S  A  V  S  F  P  P

1621  CGCAAGCAGAGACGUAGACGCAGGAGCAGGAGGACCGAAUACUGACUAACCGGGGUAGGU  1680
      R  K  Q  R  R  R  R  R  S  R  R  T  E  Y  *  L  T  G  V  G

1681  GGGUACAUAUUUUCGACGGACACAGGCCCUGGGCACUUGCAAAAGAAGUCCGUUCUGCAG  1740
      G  Y  I  F  S  T  D  T  G  P  G  H  L  Q  K  K  S  V  L  Q

1741  AACCAGCUUACAGAACCGACCUUGGAGCGCAAUGUUCUGGAAAGAAUCUACGCCCCGGUG  1800
      N  Q  L  T  E  P  T  L  E  R  N  V  L  E  R  I  Y  A  P  V

1801  CUCGACACGUCGAAAGAGGAACAGCUCAAACUCAGGUACCAGAUGAUGCCCACCGAAGCC  1860
      L  D  T  S  K  E  E  Q  L  K  L  R  Y  Q  M  M  P  T  E  A

1861  AACAAAAGCAGGUACCAGUCUAGAAAAGUAGAAAAUCAGAAAGCCAUAACCACUGAGCGA  1920
      N  K  S  R  Y  Q  S  R  K  V  E  N  Q  K  A  I  T  T  E  R

1921  CUGCUUUCAGGGCUACGACUGUAUAACUCUGCCACAGAUCAGCCAGAAUGCUAUAAGAUC  1980
      L  L  S  G  L  R  L  Y  N  S  A  T  D  Q  P  E  C  Y  K  I

1981  ACCUACCCGAAACCAUCGUAUUCCAGCAGUGUACCGGCGAACUACUCUGACCCAAAGUUU  2040
      T  Y  P  K  P  S  Y  S  S  S  V  P  A  N  Y  S  D  P  K  F

2041  GCUGUAGCUGUUUGCAACAACUAUCUGCAUGAGAAUUACCCGACGGUAGCAUCUUAUCAG  2100
      A  V  A  V  C  N  N  Y  L  H  E  N  Y  P  T  V  A  S  Y  Q

2101  AUCACCGACGAGUACGAUGCUUACUUGGAUAUGGUAGACGGGACAGUCGCUUGUCUAGAU  2160
      I  T  D  E  Y  D  A  Y  L  D  M  V  D  G  T  V  A  C  L  D

2161  ACUGCAACUUUUUGCCCCGCCAAGCUUAGAAGUUACCCGAAAAGACACGAGUAUAGAGCC  2220
      T  A  T  F  C  P  A  K  L  R  S  Y  P  K  R  H  E  Y  R  A

2221  CCAAACAUCCGCAGUGCGGUUCCAUCAGCGAUGCAGAACACGUUGCAAAACGUGCUCAUU  2280
      P  N  I  R  S  A  V  P  S  A  M  Q  N  T  L  Q  N  V  L  I

2281  GCCGCGACUAAAAGAAACUGCAACGUCACACAAAUGCGUGAAUUGCCAACACUGGACUCA  2340
      A  A  T  K  R  N  C  N  V  T  Q  M  R  E  L  P  T  L  D  S

2341  GCGACAUUCAACGUUGAAUGCUUUCGAAAAUAUGCAUGUAAUGACGAGUAUUGGGAGGAG  2400
      A  T  F  N  V  E  C  F  R  K  Y  A  C  N  D  E  Y  W  E  E

2401  UUUGCCCGAAAGCCAAUUAGGAUCACUACUGAGUUCGUUACCGCAUACGUGGCCAGACUG  2460
      F  A  R  K  P  I  R  I  T  T  E  F  V  T  A  Y  V  A  R  L

2461  AAAGGCCCUAAGGCCGCCGCACUGUUCGCAAAGACGCAUAAUUUGGUCCCAUUGCAAGAA  2520
      K  G  P  K  A  A  A  L  F  A  K  T  H  N  L  V  P  L  Q  E
```

Figure 12b. See legend on the last page of this sequence.

```
2521  GUGCCUAUGGAUAGGUUCGUCAUGGACAUGAAAAGAGACGUGAAAGUUACACCUGGCACG  2580
      V  P  M  D  R  F  V  M  D  M  K  R  D  V  K  V  T  P  G  T

2581  AAACACACAGAAGAAAGACCGAAAGUACAAGUGAUACAAGCCGCAGAACCCCUGGCGACC  2640
      K  H  T  E  E  R  P  K  V  Q  V  I  Q  A  A  E  P  L  A  T

2641  GCUUACCUGUGCGGGAUCCACCGGGAGUUAGUGCGCAGGCUUACAGCCGUCUUGCUACCC  2700
      A  Y  L  C  G  I  H  R  E  L  V  R  R  L  T  A  V  L  L  P

2701  AACAUUCACACGCUUUUUGACAUGUCGGCGGAGGACUUUGAUGCAAUCAUAGCAGAACAC  2760
      N  I  H  T  L  F  D  M  S  A  E  D  F  D  A  I  I  A  E  H

2761  UUCAAGCAAGGUGACCCGGUACUGGAGACGGAUAUCGCCUCGUUCGACAAAAGCCAAGAC  2820
      F  K  Q  G  D  P  V  L  E  T  D  I  A  S  F  D  K  S  Q  D

2821  GACGCUAUGGCGUUAACUGGCCUGAUGAUCUUGGAAGACCUGGGUGUGGACCAACCACUA  2880
      D  A  M  A  L  T  G  L  M  I  L  E  D  L  G  V  D  Q  P  L

2881  CUCGACUUGAUCGAGUGCGCCUUUGGAGAAAUAUCAUCCACCCAUCUGCCCACGGGUACC  2940
      L  D  L  I  E  C  A  F  G  E  I  S  S  T  H  L  P  T  G  T

2941  CGUUUCAAAUUCGGGGCGAUGAUGAAAUCCGGAAUGUUCCUCACGCUCUUUGUCAACACA  3000
      R  F  K  F  G  A  M  M  K  S  G  M  F  L  T  L  F  V  N  T

3001  GUUCUGAAUGUCGUUAUCGCCAGCAGAGUAUUGGAGGAGCGGCUUAAAACGUCCAAAUGU  3060
      V  L  N  V  V  I  A  S  R  V  L  E  E  R  L  K  T  S  K  C

3061  GCAGCAUUUAUCGGCGACGACAACAUCAUACACGGAGUAGUAUCUGACAAAGAAAUGGCU  3120
      A  A  F  I  G  D  D  N  I  I  H  G  V  V  S  D  K  E  M  A

3121  GAGAGGUGUGCCACCUGGCUCAACAUGGAGGUUAAGAUCAUUGACGCAGUCAUCGGCGAG  3180
      E  R  C  A  T  W  L  N  M  E  V  K  I  I  D  A  V  I  G  E

3181  AGACCGCCUUACUUCUGCGGUGGAUUCAUCUUGCAAGAUUCGGUUACCUCCACAGCGUGU  3240
      R  P  P  Y  F  C  G  G  F  I  L  Q  D  S  V  T  S  T  A  C

3241  CGCGUGGCGGACCCCUUGAAAAGGCUGUUUAAGUUGGGUAAACCGCUCCCAGCCGACGAC  3300
      R  V  A  D  P  L  K  R  L  F  K  L  G  K  P  L  P  A  D  D

3301  GAGCAAGACGAAGACAGAAGACGCGCUCUGCUAGAUGAAACAAAGGCGUGGUUUAGAGUA  3360
      E  Q  D  E  D  R  R  R  A  L  L  D  E  T  K  A  W  F  R  V

3361  GGUAUAACAGACACCUUAGCAGUGGCCGUGGCAACUCGGUAUGAGGUAGACAACAUCACA  3420
      G  I  T  D  T  L  A  V  A  V  A  T  R  Y  E  V  D  N  I  T

3421  CCUGUCCUGCUGGCAUUGAGAACUUUUGCCCAGAGCAAAAGAGCAUUUCAAGCCAUCAGA  3480
      P  V  L  L  A  L  R  T  F  A  Q  S  K  R  A  F  Q  A  I  R

3481  GGGGAAAUAAAGCAUCUCUACGGUGGUCCUAAAUAG  3516
      G  E  I  K  H  L  Y  G  G  P  K  *
```

Figure 12c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 for the Girdwood South African strain of Sindbis virus isolated in 1963.

viral disease. These viruses are all closely related, exhibiting 90% or greater amino acid sequence identity in the conserved region of nsP3 or in nsP4. Conclusions as to sequence relationships are similar to conclusions drawn from the analysis of the 3' NTR.

CONCLUSIONS

We have identified an important antigenic epitope present in E2 of the alphaviruses. This epitope, located in whole or in part within the domain of E2 between residues 170 and 220, depending upon the antibody, is clearly of major importance for the neutralization of the virus infectivity and thus for vaccine design.

We have established the relationships between many of the Sindbis-like alphaviruses. The Sindbis-like viruses, which are found throughout the Old World from Northern Europe to Africa, India, the Philippines and the Australasian region including New Guinea, are a clearly identifiable group of viruses. They share a minimum of 80% amino acid sequence identity in the nonstructural proteins and possess a characteristic and conserved 3' NTR. Virulent strains exist that can cause significant disease in man, and the relationship of the virulent strains to avirulent strains has been established. It is of considerable interest that viruses belonging to this group coexist in many parts of the world with other alphaviruses that are demonstrably different in their epidemiology, serology, organization of the 3' NTR, and evolutionary history, even though many of these non-Sindbis alphaviruses cause diseases very similar to those caused by the virulent Sindbis-like viruses.

We found that a strain of Sindbis virus from Northern Europe that causes Ockelbo disease in Sweden, Pogosta disease in Finland, or Karelian fever in

Russia, a disease characterized by a polyarthritis whose symptoms can persist for months or years, are very closely related to pathogenic strains of Sindbis virus isolated from South Africa. We concluded that a South African strain of Sindbis was introduced into Northern Europe, probably in the 1960s, where it continues to cause epidemics of a significant human disease (Shirako et al., 1991).

We have shown that Aura virus is a New World representative of the Sindbis viruses. Further analysis is required to determine whether it is one of the parents of Western equine encephalitis virus, but the hypothesis that Western equine encephalitis virus is a virus that emerged from a recombination event has received further support from these studies.

We have also shown that high throughput automated DNA sequencing is ideally suited to the rapid analysis of an RNA virus family such as the alphaviruses. These procedures are rapid and generate large amounts of useful information very quickly. Such procedures would be very useful in defining the origin and spread of an epidemic virus.

REFERENCES

Bell, J. R., Hunkapiller, M. W., Hood, L. E., and Strauss, J. H. (1978). Amino-terminal sequence analysis of the structural proteins of Sindbis virus. *Proc. Natl. Acad. Sci. USA* **75,** 2722-2726.

Causey, O. R., Casals, J., Shope, R. E., and Udomsakdi, S. (1963). Aura and Una, two new group A arthropod-borne viruses. *Am. J. Trop. Med. Hyg.* **12,** 777-781.

Davis, N. L., Fuller, F. J., Dougherty, W. G., Olmsted, R. A., and Johnston, R. E. (1986). A single nucleotide change in the E2 glycoprotein gene of Sindbis virus affects penetration rate in cell culture and virulence in neonatal mice. *Proc. Natl. Acad. Sci. USA* **83,** 6771-6775.

Griffin, D. E. (1986). Alphavirus pathogenesis and immunity. *In* "The Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Ed.), pp. 209-250. Plenum Publishing Corp., New York.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25,** 263-269.

Hahn, C. S., Lustig, S., Strauss, E. G., and Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc. Natl. Acad Sci. USA* **85,** 5997-6001.

Hahn, C. S., Strauss, E. G., and Strauss, J. H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid autoprotease. *Proc. Natl. Acad. Sci. USA* **82,** 4648-4652.

Hsu, M. T., Kung, H. J., and Davidson. N. (1973). An electron microscope study of Sindbis virus RNA. *Cold Spring Harbor Symp. Quant. Biol.* **38,** 943-950.

Kuhn, R. J., Hong, Z., and Strauss, J. H. (1990). Mutagenesis of the 3' nontranslated region of Sindbis virus RNA. *J. Virol.* **64,** 1465-1476.

Levinson, R., Strauss, J. H., and Strauss, E. G. (1990). Determination of the complete nucleotide sequence of the genomic RNA of O'Nyong-nyong virus and its use in the construction of phylogenetic trees. *Virology* **175,** 110-123.

Lustig, S., Jackson, A., Hahn, C. S., Griffin, D. E., Strauss, E. G., and Strauss, J. H. (1988). Molecular basis of Sindbis virus neurovirulence in mice. *J. Virol.* **62,** 2329-2336.

Lvov, D. K., Berezina, L. K., Yakovlev, B. I., Aristova, V. A., Gushchina, E. L., Lvov, S. D., Myasnikova, I. A., Skvortsova, T. M., Gromashevsky, V. L., Gushchin, B. V., Sidorova, G. A., Klimenko, S. M., Khutoretskaya, N. V., and Khizhnyakova, T. M. (1984). Isolation of Karelian fever agent from *Aedes communis* mosquitoes. *Lancet* **2,** 399-400.

Lvov, D. K., Vladimirtseva, E. A., Butenko, A. M., Karabatsos, N., Trent, D. W., and Calisher, C. H. (1988). Identity of Karelian fever and Ockelbo viruses determined by serum dilution-plaque reduction neutralization tests and oligonucleotide mapping. *Am J Trop Med Hyg* **39,** 607-610.

Malherbe, H., Strickland-Cholmley, M., and Jackson, A. L. (1963). Sindbis virus infection in man: Report of a case with recovery of virus from skin lesions. *S. Afr. Med. J.* **37,** 547-552.

Niklasson, B., Aspmark, A., LeDuck, J. W., Gargan, T. P., Ennis, W. A., Tesh, R. B., and A. J. Main, J. (1984). Association of a Sindbis-like virus with Ockelbo disease in Sweden. *Am. J. Trop. Med. Hyg.* **33,** 1212-1217.

Olmsted, R. A., Meyer, W. J., and Johnston, R. E. (1986). Characterization of Sindbis virus epitopes important for penetration in cell culture and pathogenesis in animals. *Virology* **148,** 245-254.

Ou, J.-H., Strauss, E. G., and Strauss, J. H. (1981). Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. *Virology* **109,** 281-289.

Peters, C. J., and Dalrymple, J. M. (1990). Alphaviruses. *In* "Virology" (B. N. Fields and D. M. Knipe, Ed.), pp. 713-761. Raven Press, New York.

Pierce, J. S., Strauss, E. G., and Strauss, J. H. (1974). Effect of ionic strength on the binding of Sindbis virus to chick cells. *J. Virol.* **13,** 1030-1036.

Rice, C., M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L., and Strauss, J. H. (1985). Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* **229,** 726-733.

Rice, C. M., and Strauss, J. H. (1981). Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus. *J. Mol. Biol.* **150,** 315-340.

Russell, D. L., Dalrymple, J. M., and Johnston, R. E. (1989). Sindbis virus mutations which coordinately affect glycoprotein processing, penetration and virulence in mice. *J. Virol.* **63,** 1619-1629.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Shah, K. V., Johnson, H. N., Rao, T. R., Rajagopalan, P. K., and Lamba, B. S. (1960). Isolation of five strains of Sindbis virus in India. *Ind. J. Med. Res.* **48,** 300-308.

Shirako, Y., Niklasson, B., Dalrymple, J. M., Strauss, E. G., and Strauss, J. H. (1991). Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology* **182,** 753-764.

Stec, D. S., Waddell, A., Schmaljohn, C. S., Cole, G. A., and Schmaljohn, A. L. (1986). Antibody-selected variation and reversion in Sindbis Virus neutralization epitopes. *J. Virol.* **57,** 715-720.

Strauss, E. G., Rice, C. M., and Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133,** 92-110.

Strauss, E. G., Stec, D. S., Schmaljohn, A. L., and Strauss, J. H. (1991). Identification of antigenically important domains in the glycoproteins of Sindbis virus by analysis of antibody escape variants. *J. Virol.* **65,** 4654-4664.

Taylor, R. M., Hurlbut, H. S., Work, T. H., Kingston, J. R., and Frothingham, T. E. (1955). Sindbis virus: A newly recognized arthropod-transmitted virus. *Am. J. Trop. Med. Hyg.* **4,** 844-862.

Tucker, P. C., and Griffin, D. E. (1991). Mechanism of altered Sindbis virus neurovirulence associated with a single-amino-acid change in the E2 glycoprotein. *J. Virol.* **65,** 1551-1557.

Ubol, S., and Griffin, D. E. (1991). Identification of a putative alphavirus receptor on mouse neural cells. *J. Virol.* **65,** 6913-6921.

Wang, K.-S., Schmaljohn, A. L., Kuhn, R. J., and Strauss, J. H. (1991). Antiidiotypic antibodies as probes for the Sindbis virus receptor. *Virology* **181,** 694-702.

Wang, K.-S., and Strauss, J. H. (1991). Use of a λgt11 expression library to localize a neutralizing antibody-binding site in glycoprotein E2 of Sindbis virus. *J. Virol.* **65,** 7037-7040.

Young, R. A., and Davis, R. W. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA,* **80,** 1194-1198.